



中国科学院自动化研究所
模式识别实验室
New Laboratory of Pattern Recognition



多模态人工智能系统
全国重点实验室
State Key Laboratory of
Multimodal Artificial Intelligence Systems



中国科学院自动化研究所
Institute of Automation
Chinese Academy of Sciences



[AI for Science]

分子表征学习与构效谱关系建模

王亮

中国科学院自动化研究所 – 多模态人工智能系统全国重点实验室
新加坡国立大学 – 计算机学院

目录

1 / 研究背景

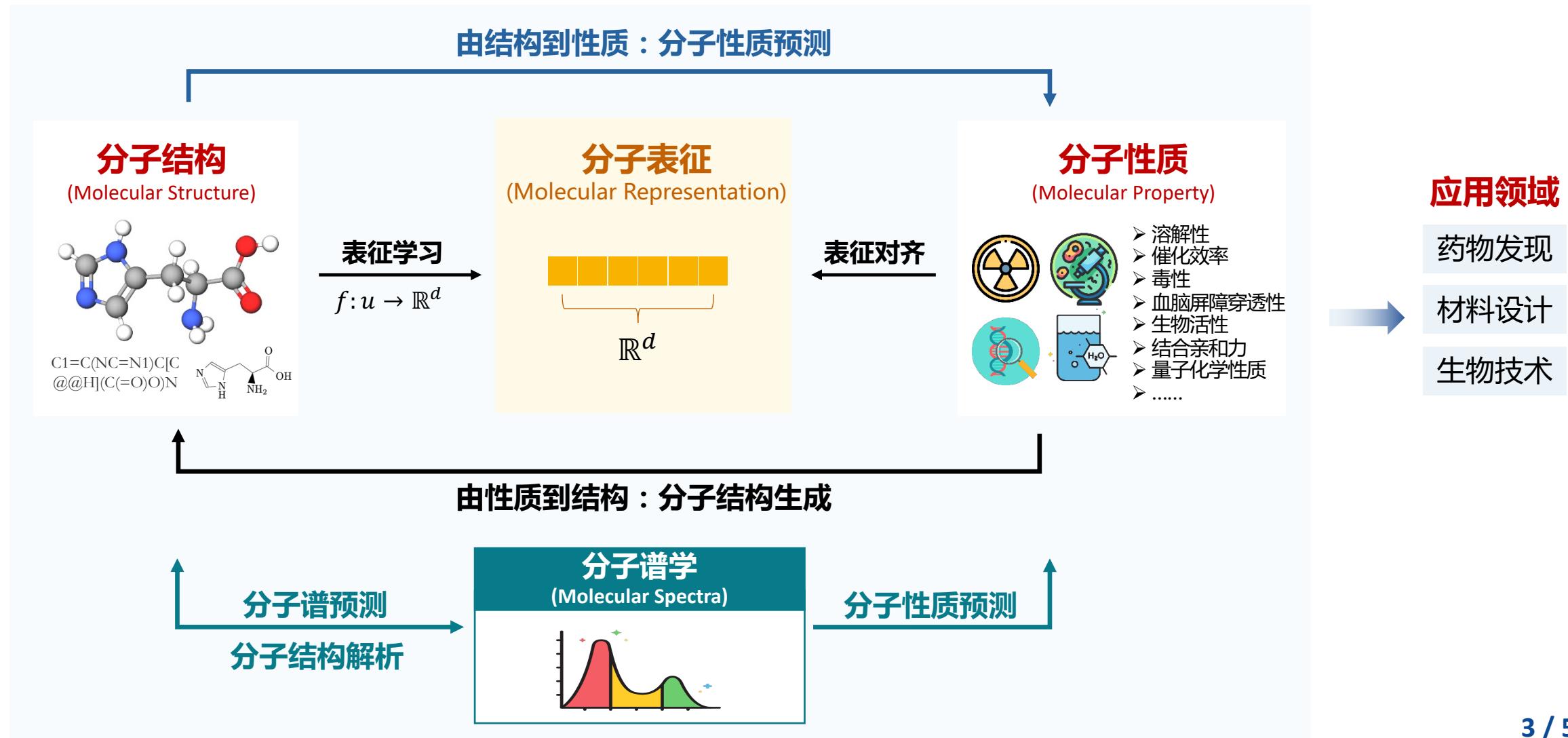
2 / 理解：分子表征学习

3 / 预测：分子性质预测

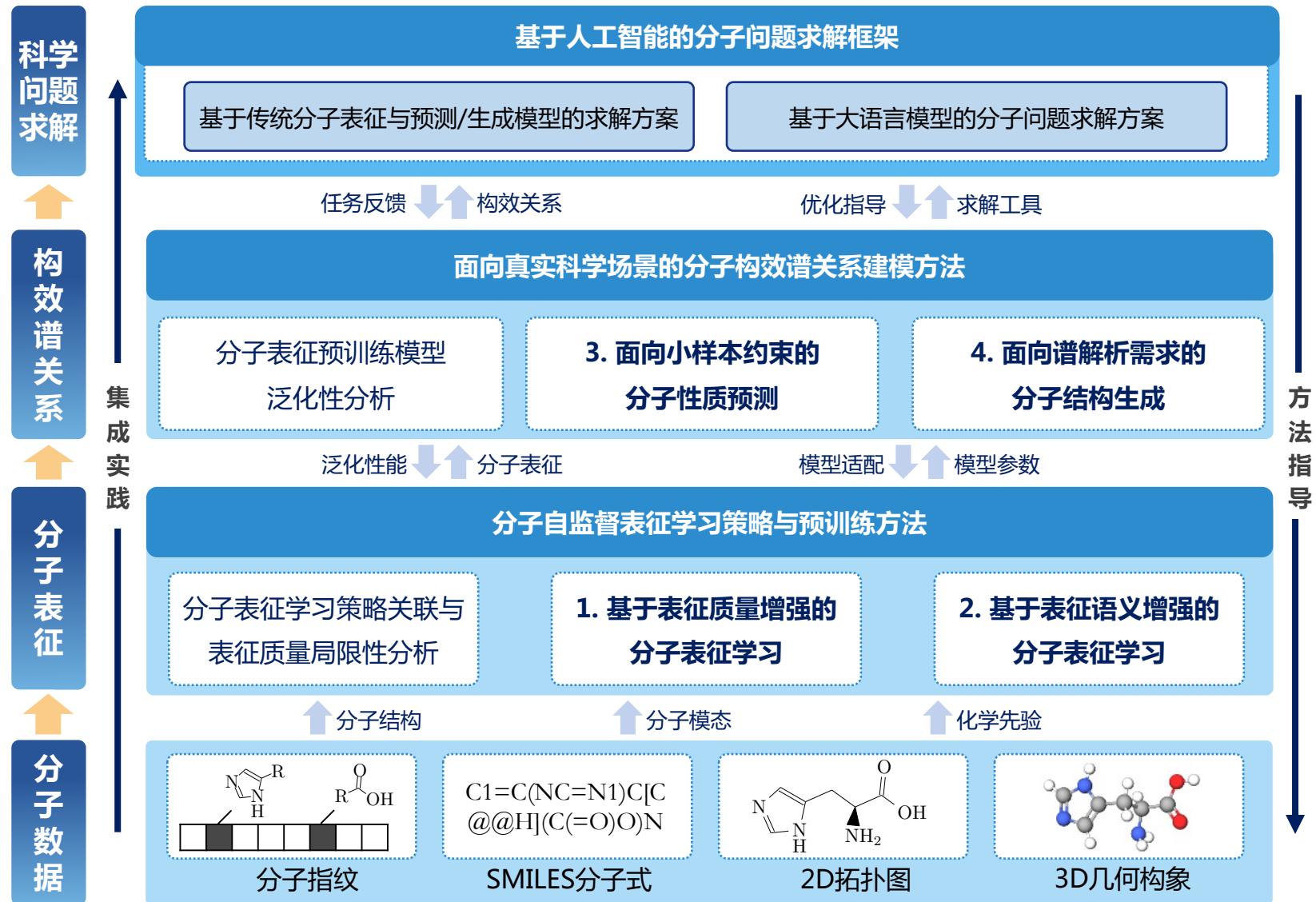
4 / 生成：分子结构解析

研究背景

分子表征学习与构效谱关系建模



研究背景



研究成果

分享论文

表征质量增强的分子表征预训练

- **Liang Wang**, Xiang Tao, Qiang Liu, Shu Wu, and Liang Wang, "Rethinking Graph Masked Autoencoders through Alignment and Uniformity." **AAAI 2024**.

表征语义增强的分子表征预训练

- **Liang Wang**, Shaozhen Liu, Yu Rong, Deli Zhao, Qiang Liu, Shu Wu, Liang Wang, "MolSpectra: Pre-training 3D Molecular Representation with Multi-modal Energy Spectra" **ICLR 2025**.

小样本约束的分子性质预测

- **Liang Wang**, Qiang Liu, Shaozhen Liu, Xin Sun, Shu Wu, Liang Wang, "Pin-Tuning: Parameter-Efficient In-Context Tuning for Few-shot Molecular Property Prediction" **NeurIPS 2024**.

基于谱条件的分子结构生成

- **Liang Wang**, Yu Rong, Tingyang Xu, Zhenyi Zhong, Zhiyuan Liu, Pengju Wang, Deli Zhao, Qiang Liu, Shu Wu, Liang Wang, "DiffSpectra: Molecular Structure Elucidation from Spectra using Diffusion Models" **arXiv 2025**.

目录

1 / 研究背景

2 / 理解：分子表征学习

3 / 预测：分子性质预测

4 / 生成：分子结构解析



中国科学院自动化研究所
模式识别实验室
New Laboratory of Pattern Recognition



多模态人工智能系统
全国重点实验室
State Key Laboratory of
Multimodal Artificial Intelligence Systems



中国科学院自动化研究所
Institute of Automation
Chinese Academy of Sciences



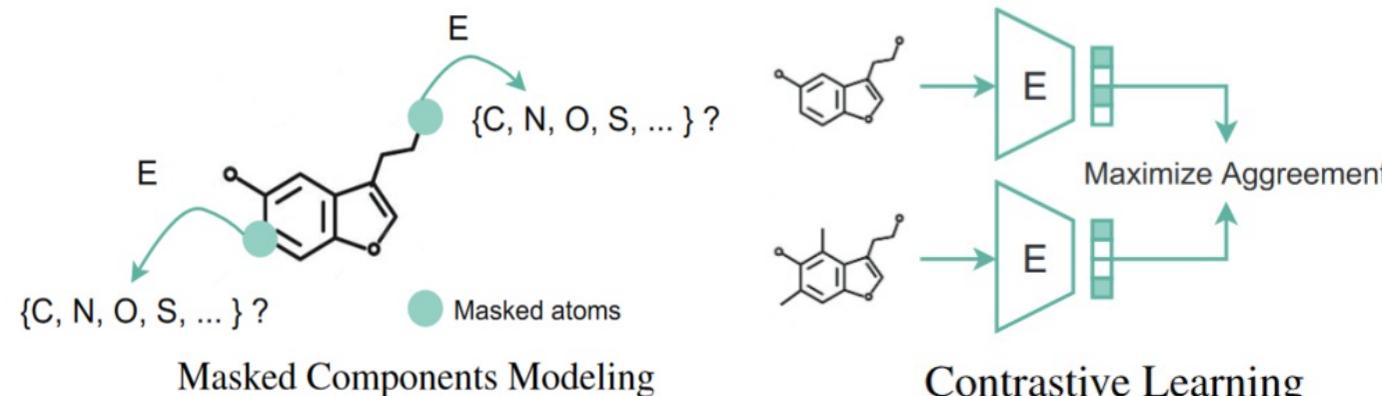
[AAAI 2024] Rethinking Graph Masked Autoencoders through Alignment and Uniformity

Liang Wang^{1,2}, Xiang Tao^{1,2}, Qiang Liu^{1,2}, Shu Wu^{1,2}, Liang Wang^{1,2}

¹Institute of Automation, Chinese Academy of Sciences

²University of Chinese Academy of Sciences

工作一：(表征质量增强) 基于表征一致性与均匀性增强的分子表征学习



重构式方法和对比式方法是完全不同的两类方法，还是存在内在关联？

Are GraphMAE and GCL completely different methods, or do they share any commonality?

工作一：(表征质量增强) 基于表征一致性与均匀性增强的分子表征学习

理论分析

Assumption 4.1. For any graph decoder g , we assume the existence of a pseudo-inverse graph encoder f_g such that the resulting pseudo graph autoencoder $h_g = g \circ f_g$ satisfies $\mathbb{E}_x \|h_g(\mathbf{x}) - \mathbf{x}\|^2 \leq \varepsilon$, where \mathbf{x} represents the feature of masked node $v \in \tilde{\mathcal{V}}$.

Theorem 4.2. Under Assumption 4.1, the SCE loss in Eq. (2) can be lower bounded by a pretext loss:

$$\begin{aligned} \mathcal{L}_{\text{SCE}}(h) &\geq \frac{\gamma}{2} \mathcal{L}_{\text{Pretext}}(h) - \frac{\gamma}{2} \varepsilon + \text{const}, \\ \text{where } \mathcal{L}_{\text{Pretext}}(h) &= -\mathbb{E}_{v_i \in \tilde{\mathcal{V}}} h_g(\mathbf{x}_i)^\top h(c_i). \end{aligned} \quad (6)$$

Definition 4.3. (Context-Level Alignment Loss) The alignment loss for positive context pairs (c, c^+) is defined as:

$$\mathcal{L}_{\text{Align}}^c(h) = -\mathbb{E}_{(c, c^+) \sim p_{\text{pos}}^c} h(c)^\top h(c^+). \quad (7)$$

Theorem 4.4. The pretext loss in Eq. (6) can be lower bounded by the context-level alignment loss in Eq. (7):

$$\mathcal{L}_{\text{Pretext}}(h) \geq \frac{1}{2} \mathcal{L}_{\text{Align}}^c(h) + \text{const}. \quad (8)$$

Theorem 4.5. Under Assumption 4.1, GraphMAE's node-level reconstruction loss in Eq. (2) can be lower bounded by the context-level alignment loss in Eq. (7):

$$\begin{aligned} \mathcal{L}_{\text{SCE}}(h) &\geq \frac{\gamma}{4} \mathcal{L}_{\text{Align}}^c(h) - \frac{\gamma}{2} \varepsilon + \text{const} \\ &= -\frac{\gamma}{4} \mathbb{E}_{c, c^+} h(c)^\top h(c^+) - \frac{\gamma}{2} \varepsilon + \text{const}. \end{aligned} \quad (10)$$

Proof Sketch

$$\begin{aligned} \mathcal{L}_{\text{SCE}} &= \mathbb{E}_{v_i \in \tilde{\mathcal{V}}} (1 - \mathbf{x}_i^\top h(c_i))^{\gamma} \\ &\geq \mathbb{E}_{v_i \in \tilde{\mathcal{V}}} (1 - \gamma \mathbf{x}_i^\top h(c_i)) \quad (\text{Bernoulli's inequality}) \\ &= \mathbb{E}_{v_i \in \tilde{\mathcal{V}}} (1 - \gamma(1 - \frac{1}{2} \|\mathbf{x}_i - h(c_i)\|^2)) \quad (\text{features are normalized}) \\ &= 1 - \gamma + \frac{\gamma}{2} \mathbb{E}_{v_i \in \tilde{\mathcal{V}}} \|\mathbf{x}_i - h(c_i)\|^2 \\ &= 1 - \gamma + \frac{\gamma}{2} \mathbb{E}_{v_i \in \tilde{\mathcal{V}}} (\|\mathbf{x}_i - h(c_i)\|^2 + \varepsilon) - \frac{\gamma}{2} \varepsilon \quad (\text{Assumption 4.1}) \\ &\geq 1 - \gamma + \frac{\gamma}{2} \mathbb{E}_{v_i \in \tilde{\mathcal{V}}} (\|\mathbf{x}_i - h(c_i)\|^2 + \|h_g(\mathbf{x}_i) - \mathbf{x}_i\|^2) - \frac{\gamma}{2} \varepsilon. \\ &\geq 1 - \gamma + \frac{\gamma}{4} \mathbb{E}_{v_i \in \tilde{\mathcal{V}}} \|h_g(\mathbf{x}_i) - h(c_i)\|^2 - \frac{\gamma}{2} \varepsilon \\ &= 1 - \gamma + \frac{\gamma}{4} \mathbb{E}_{v_i \in \tilde{\mathcal{V}}} (2 - 2h_g(\mathbf{x}_i)^\top h(c_i)) - \frac{\gamma}{2} \varepsilon \\ &= -\frac{\gamma}{2} \mathbb{E}_{v_i \in \tilde{\mathcal{V}}} h_g(\mathbf{x}_i)^\top h(c_i) - \frac{\gamma}{2} \varepsilon + 1 - \frac{\gamma}{2} \\ &= \frac{\gamma}{2} \mathcal{L}_{\text{Pretext}}(h) - \frac{\gamma}{2} \varepsilon + \text{const}. \end{aligned}$$

Proof Sketch

$$\begin{aligned} \mathcal{L}_{\text{Pretext}}(h) &= -\text{tr}(\mathbf{H}_g^\top \tilde{\mathbf{A}}_{\text{CF}} \mathbf{H}) \\ &\geq -\frac{1}{2} \left(\|\mathbf{H}_g\|_{\text{F}}^2 + \|\tilde{\mathbf{A}}_{\text{CF}} \mathbf{H}\|_{\text{F}}^2 \right) \quad (\text{tr}(\mathbf{AB}) \leq \frac{1}{2} (\|\mathbf{A}\|_{\text{F}}^2 + \|\mathbf{B}\|_{\text{F}}^2)) \\ &= -\frac{1}{2} \text{tr}(\tilde{\mathbf{A}}_{\text{CF}}^\top \tilde{\mathbf{A}}_{\text{CF}} \mathbf{H} \mathbf{H}^\top) - \frac{1}{2} \quad (\|\mathbf{H}_g\|_{\text{F}}^2 = \sum_{f_j} d_{f_j} \|h_g(f_j)\|^2 = 1) \\ &= -\frac{1}{2} \sum_{c, c^+} \sum_{f_j} \frac{w_{c, f_j} w_{c^+, f_j}}{d_{f_j}} h(c)^\top h(c^+) - \frac{1}{2} \\ &= -\frac{1}{2} \sum_{c, c^+} (\mathbf{A}_C)_{c, c^+} h(c)^\top h(c^+) - \frac{1}{2} \\ &= \frac{1}{2} \mathcal{L}_{\text{align}}^c(h) - \frac{1}{2}, \end{aligned}$$

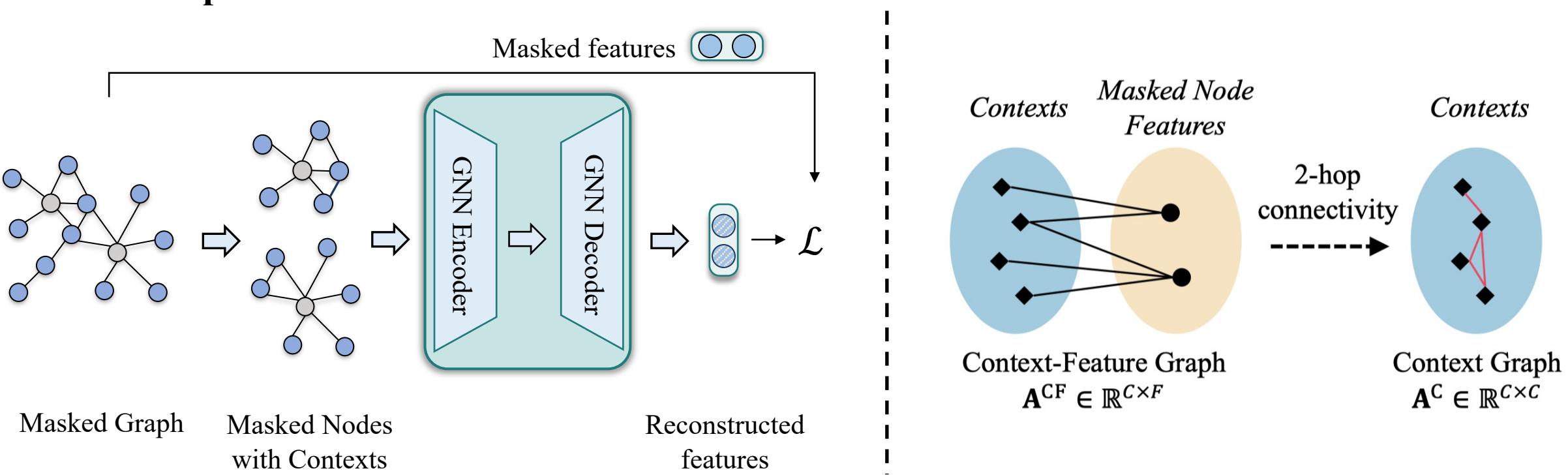
理论结果: 子图 (子结构)
级别的对比学习目标，给出了节点 (原子) 级别重构学习目标的下界

工作一：(表征质量增强) 基于表征一致性与均匀性增强的分子表征学习

理论分析

Theoretical Result: 基于节点（原子）级别的特征重构式方法，隐式地执行了子图（子结构）级别的对比学习

Intuitive Explanation:

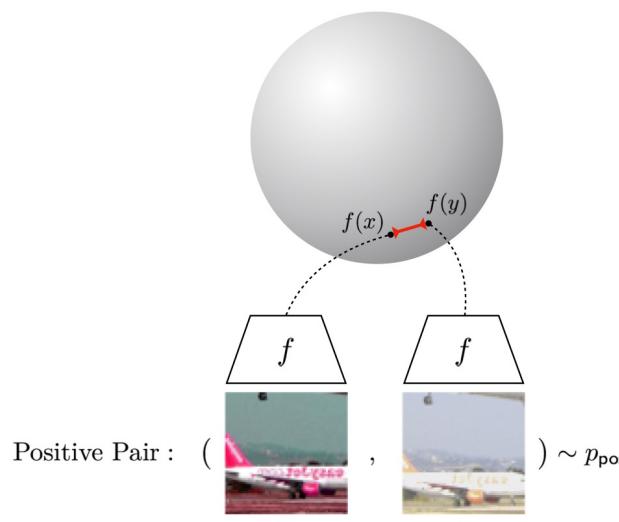


工作一：(表征质量增强) 基于表征一致性与均匀性增强的分子表征学习

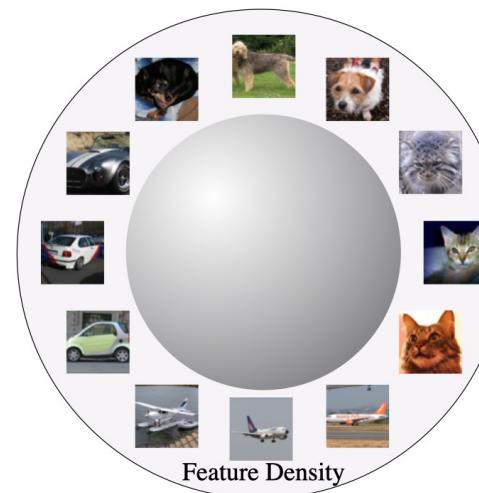
理论分析

表征质量评价：

- 表征一致性 (Representation Alignment) ?
- 表征均匀性 (Representation Uniformity) ?



Alignment: Similar samples have similar features.
(Figure inspired by Tian et al. (2019).)



Uniformity: Preserve maximal information.

Note:

- **Alignment (一致性)** refers to the concentration of samples from the same class within the same region of the hypersphere.
- **Uniformity (均匀性)** refers to the uniform distribution of all samples on the hypersphere.

工作一：(表征质量增强) 基于表征一致性与均匀性增强的分子表征学习

理论分析

Limitations of GraphMAE:

- Alignment performance is still restricted by the mask distribution, which is decided by the masking strategy.

$$\mathcal{L}_{\text{SCE}} = \mathbb{E}_{v_i \in \widetilde{\mathcal{V}}} \left(1 - x_i^T \cdot g(f(c_i)) \right)^\gamma, \gamma \geq 1,$$

$$\mathcal{L}_{\text{Align}}^c(h) = - \mathbb{E}_{c, c^+} h(c)^\top h(c^+).$$

- Uniformity performance is **not strictly guaranteed**.

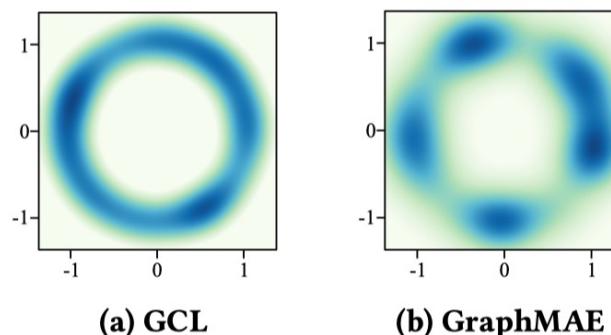
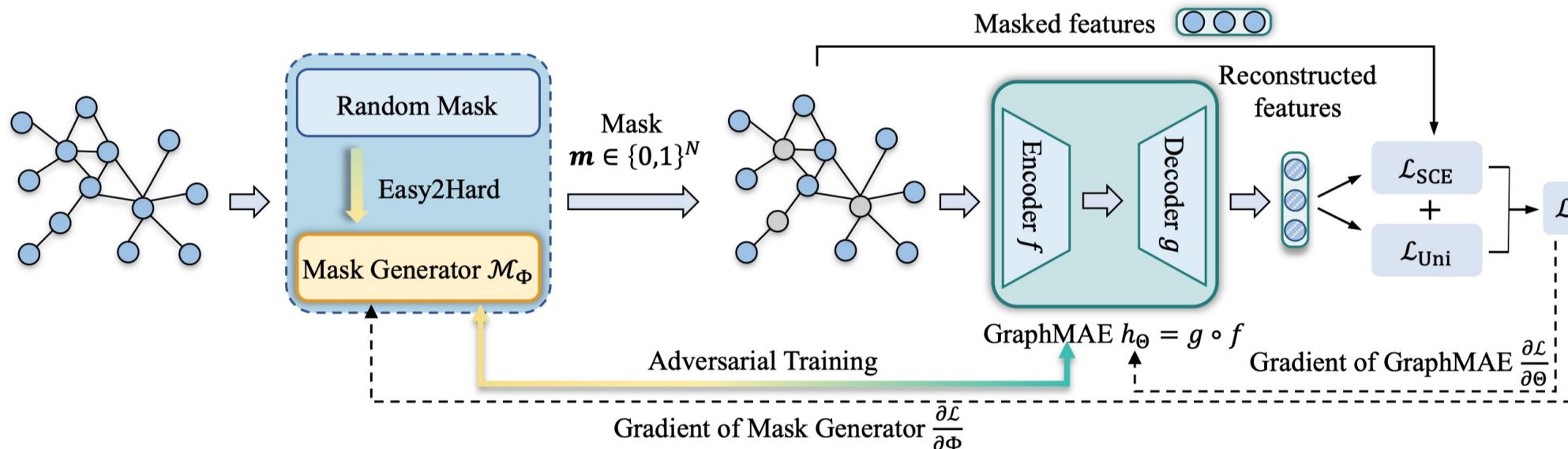


Figure 1: Distribution of nodes representations on the unit hypersphere learned by GCL (taking GRACE [52] as an example) and GraphMAE [7]. The representations learned by GCL is more uniformly distributed than GraphMAE.

工作一：(表征质量增强) 基于表征一致性与均匀性增强的分子表征学习

Alignment-Uniformity Enhanced Graph Masked Autoencoders



Alignment Enhancement

- Adversarial Masking

$$\Phi^\star = \arg \max_{\Phi} (\mathcal{L}_{\text{SCE}}(\mathcal{G}; \Theta, \Phi) - \lambda_1 \sin(\frac{\pi}{N} \sum_{i=1}^N m_i)^{-1}),$$

$$\Theta^\star = \arg \min_{\Theta} (\mathcal{L}_{\text{SCE}}(\mathcal{G}; \Theta, \Phi) + (1 - \alpha_{\text{adv}})\lambda_2 \mathcal{L}_{\text{Uni}}(\mathcal{G}; \Theta)),$$

- Easy-to-Hard Masking

$$prob(t) = (1 - \alpha_{\text{adv}}(t)) \cdot prob_{\text{rand}} + \alpha_{\text{adv}}(t) \cdot prob_{\text{adv}}(t),$$

$$\alpha_{\text{adv}}(t) = \alpha_0 + \Delta\alpha(t) = \alpha_0 + (\frac{t}{T})^\eta \cdot (\alpha_T - \alpha_0),$$

Uniformity Enhancement

- Explicit Uniformity Regularizer

$$\mathcal{L}_{\text{Uni}} = \log \mathbb{E}_{(z_i, z_j) \sim p_{\text{data}}} e^{-t \|z_i - z_j\|^2},$$

工作一：(表征质量增强) 基于表征一致性与均匀性增强的分子表征学习

实验结果

节点分类和图分类实验结果

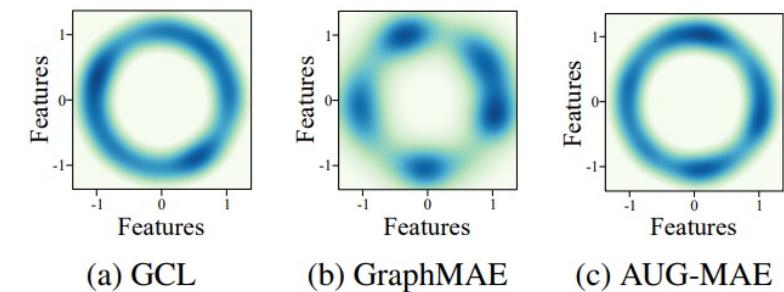
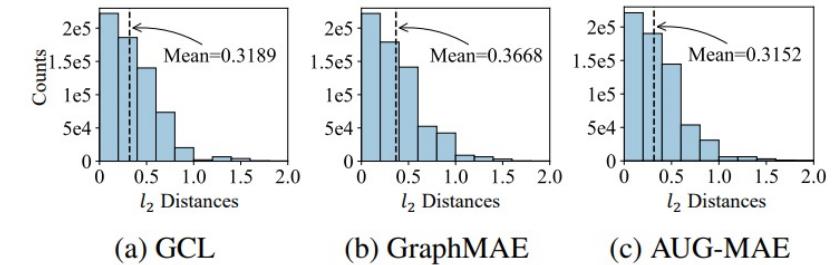
	Method	Cora	CiteSeer	PubMed	Ogbn-arxiv	PPI	Reddit	Corafull	Flickr	WikiCS	A.R.
Contrastive	DGI	82.3 ± 0.6	71.8 ± 0.7	76.8 ± 0.6	70.3 ± 0.2	63.8 ± 0.2	94.0 ± 0.1	48.2 ± 0.5	45.0 ± 0.2	64.8 ± 0.6	7.89
	MVGRL	83.5 ± 0.4	<u>73.3 ± 0.5</u>	80.1 ± 0.7	-	-	-	52.6 ± 0.5	-	64.8 ± 0.7	5.20
	GRACE	81.9 ± 0.4	<u>71.2 ± 0.5</u>	80.6 ± 0.4	71.5 ± 0.1	69.7 ± 0.2	94.7 ± 0.1	45.2 ± 0.1	-	68.0 ± 0.7	6.50
	BGRL	82.7 ± 0.6	71.1 ± 0.8	79.6 ± 0.5	<u>71.6 ± 0.1</u>	73.6 ± 0.2	94.2 ± 0.1	47.4 ± 0.5	39.4 ± 0.1	65.5 ± 1.5	6.56
	InfoGCL	83.5 ± 0.3	73.5 ± 0.4	79.1 ± 0.2	-	-	-	-	-	-	4.67
	CCA-SSG	84.0 ± 0.4	73.1 ± 0.3	81.0 ± 0.4	71.2 ± 0.2	73.3 ± 0.2	95.1 ± 0.1	<u>53.5 ± 0.4</u>	49.1 ± 0.1	67.4 ± 0.9	3.89
Generative	SeeGera	82.8 ± 0.3	71.6 ± 0.2	79.2 ± 0.3	71.2 ± 0.3	73.4 ± 0.3	95.2 ± 0.2	52.0 ± 0.4	49.4 ± 0.5	65.8 ± 0.2	5.78
	MaskGAE	82.6 ± 0.3	73.1 ± 0.6	<u>81.0 ± 0.3</u>	71.2 ± 0.3	73.9 ± 0.3	95.4 ± 0.1	52.2 ± 0.1	49.1 ± 0.4	66.0 ± 0.2	4.78
	GraphMAE	84.0 ± 0.6	73.1 ± 0.4	80.9 ± 0.4	71.3 ± 0.6	<u>74.1 ± 0.4</u>	<u>95.8 ± 0.4</u>	53.3 ± 0.4	<u>49.5 ± 0.5</u>	70.6 ± 0.9	3.00
	AUG-MAE	84.3 ± 0.4	73.2 ± 0.4	81.4 ± 0.4	71.9 ± 0.2	74.3 ± 0.1	96.1 ± 0.1	57.6 ± 0.3	50.3 ± 0.2	71.7 ± 0.6	1.22

Table 1: Node classification results on benchmarks. We report Micro-F1(%) score for PPI and accuracy(%) for the other datasets. The best results are highlighted in **bold** and the runner ups are highlighted with underlines. A.R. means the average rank.

	Method	IMDB-B	IMDB-M	PROTEINS	COLLAB	MUTAG	REDDIT-B	A.R.
Contrastive	Graph2vec	71.10 ± 0.54	50.44 ± 0.87	73.30 ± 2.05	-	83.15 ± 9.25	75.78 ± 1.03	7.00
	InfoGraph	73.03 ± 0.87	49.69 ± 0.53	74.44 ± 0.31	70.65 ± 1.13	89.01 ± 1.13	82.50 ± 1.42	5.17
	GraphCL	71.14 ± 0.44	48.58 ± 0.67	74.39 ± 0.45	71.36 ± 1.15	86.80 ± 1.34	<u>89.53 ± 0.84</u>	5.83
	JOAO	70.21 ± 3.08	49.20 ± 0.77	74.55 ± 0.41	69.50 ± 0.36	87.35 ± 1.02	85.29 ± 1.35	6.33
	GCC	72.0	49.4	-	78.9	-	89.8	4.50
	MVGRL	74.20 ± 0.70	51.20 ± 0.50	-	-	89.70 ± 1.10	84.50 ± 0.60	4.00
	InfoGCL	75.10 ± 0.90	<u>51.40 ± 0.80</u>	-	80.00 ± 1.30	88.28 ± 0.98	-	2.25
Generative	GraphMAE	<u>75.30 ± 0.59</u>	51.35 ± 0.78	<u>75.30 ± 0.52</u>	<u>80.32 ± 0.42</u>	88.19 ± 1.26	87.83 ± 0.25	3.00
	AUG-MAE	75.56 ± 0.61	51.80 ± 0.86	75.83 ± 0.24	80.48 ± 0.50	91.20 ± 1.30	87.98 ± 0.43	1.83

Table 2: Graph classification results on benchmarks. We report accuracy(%) for all datasets. The best results are highlighted in **bold** and the runner ups are highlighted with underlines. A.R. means the average rank.

分子表征一致性与均匀性可视化



工作一：(表征质量增强) 基于表征一致性与均匀性增强的分子表征学习

实验结果

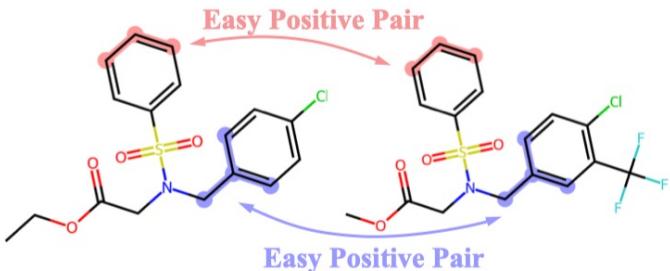
分子表征预训练和分子性质预测实验结果

	BBBP	Tox21	ToxCast	SIDER	ClinTox	MUV	HIV	BACE	Avg.
No-pretrain	65.5±1.8	74.3±0.5	63.3±1.5	57.2±0.7	58.2±2.8	71.7±2.3	75.4±1.5	70.0±2.5	67.0
ContextPred	64.3±2.8	75.7±0.7	63.9±0.6	60.9±0.6	65.9±3.8	75.8±1.7	77.3±1.0	79.6±1.2	70.4
AttrMasking	64.3±2.8	76.7±0.4	64.2±0.5	61.0±0.7	71.8±4.1	74.7±1.4	77.2±1.1	79.3±1.6	71.1
Infomax	68.8±0.8	75.3±0.5	62.7±0.4	58.4±0.8	69.9±3.0	75.3±2.5	76.0±0.7	75.9±1.6	70.3
GraphCL	69.7±0.7	73.9±0.7	62.4±0.6	60.5±0.9	76.0±2.7	69.8±2.7	<u>78.5±1.2</u>	75.4±1.4	70.8
JOAO	70.2±1.0	75.0±0.3	62.9±0.5	60.0±0.8	81.3±2.5	71.7±1.4	76.7±1.2	77.3±0.5	71.9
GraphLoG	72.5±0.8	75.7±0.5	63.5±0.7	61.2±1.1	76.7±3.3	76.0±1.1	77.8±0.8	83.5±1.2	73.4
GraphMAE	72.0±0.6	75.5±0.6	64.1±0.3	60.3±1.1	82.3±1.2	<u>76.3±2.4</u>	77.2±1.0	83.1±0.9	<u>73.8</u>
AUG-MAE	71.8±0.2	<u>76.1±0.3</u>	64.5±0.3	<u>62.5±1.0</u>	<u>81.4±1.3</u>	78.4±0.2	78.7±0.6	84.9±0.2	74.8

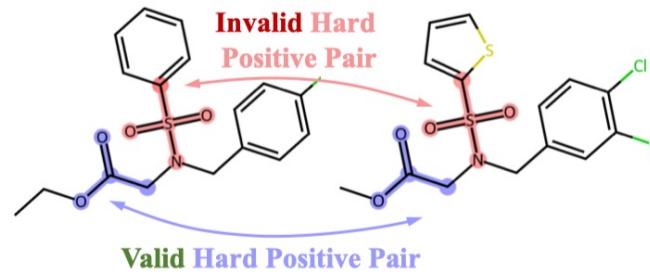
Table 3: Performance comparison on molecular property prediction benchmarks. The model is first pre-trained on ZINC15 and then finetuned on the following datasets. We reported ROC-AUC scores(%).

工作一：(表征质量增强) 基于表征一致性与均匀性增强的分子表征学习

问题：在构造难样本对的同时，也构造了无效样本对



(a) Positive pairs constructed by random masking.



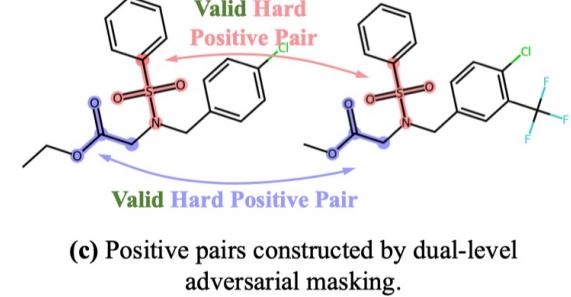
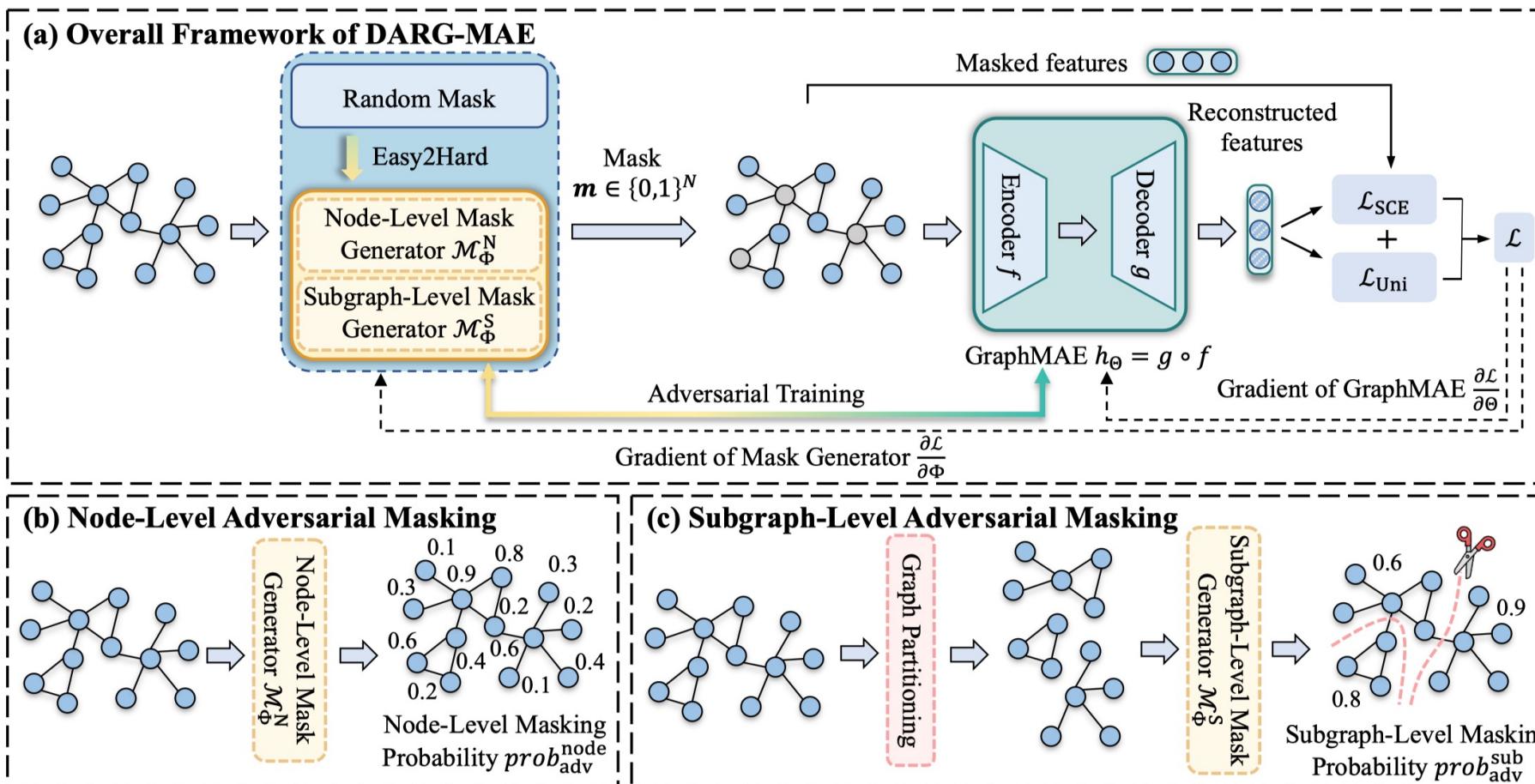
(b) Positive pairs constructed by node-level adversarial masking.

TABLE I: Comparison of properties between benzenesulfonyl and thiophenesulfonyl groups. These two fragments have many different properties, which lead to the invalid positive pair constructed at the node-level adversarial masking in Fig. 4(b).

Property		
Aromatic Ring Type	Benzene Ring	Thiophene Ring
Aromaticity	Strong	Weak
Electronic Distribution	Uniform	Non-Uniform
Melting/Boiling Points	High	Low
Chemical Stability	High	Low
Chemical Reactivity	Low	High

工作一：(表征质量增强) 基于表征一致性与均匀性增强的分子表征学习

Dual-level adversarial masking



工作一：(表征质量增强) 基于表征一致性与均匀性增强的分子表征学习

分子表征预训练和分子性质预测实验结果

TABLE VI: Performance comparison on molecular property prediction benchmarks. The model is first pre-trained on ZINC15 and then finetuned on the following datasets. We reported ROC-AUC scores(%). Avg. means the average score on all datasets.

	BBBP	Tox21	ToxCast	SIDER	ClinTox	MUV	HIV	BACE	Avg.
No-pretrain	65.5±1.8	74.3±0.5	63.3±1.5	57.2±0.7	58.2±2.8	71.7±2.3	75.4±1.5	70.0±2.5	67.0
ContextPred	64.3±2.8	75.7±0.7	63.9±0.6	60.9±0.6	65.9±3.8	75.8±1.7	77.3±1.0	79.6±1.2	70.4
DGI	68.8±0.8	75.3±0.5	62.7±0.4	58.4±0.8	69.9±3.0	75.3±2.5	76.0±0.7	75.9±1.6	70.3
GraphCL	69.7±0.7	73.9±0.7	62.4±0.6	60.5±0.9	76.0±2.7	69.8±2.7	78.5±1.2	75.4±1.4	70.8
JOAO	70.2±1.0	75.0±0.3	62.9±0.5	60.0±0.8	81.3±2.5	71.7±1.4	76.7±1.2	77.3±0.5	71.9
GraphLoG	72.5±0.8	75.7±0.5	63.5±0.7	61.2±1.1	76.7±3.3	76.0±1.1	77.8±0.8	83.5±1.2	73.4
AttrMasking	64.3±2.8	76.7±0.4	64.2±0.5	61.0±0.7	71.8±4.1	74.7±1.4	77.2±1.1	79.3±1.6	71.1
GraphMAE	72.0±0.6	75.5±0.6	64.1±0.3	60.3±1.1	82.3±1.2	76.3±2.4	77.2±1.0	83.1±0.9	73.9
DARG-MAE	73.0±0.2	76.8±0.2	64.6±0.2	62.8±0.9	84.4±0.8	78.8±0.2	79.6±0.5	86.3±0.3	75.8



中国科学院自动化研究所
模式识别实验室
New Laboratory of Pattern Recognition



多模态人工智能系统
全国重点实验室
State Key Laboratory of
Multimodal Artificial Intelligence Systems



中国科学院自动化研究所
Institute of Automation
Chinese Academy of Sciences



NUS
National University
of Singapore

達摩院
DAMO ACADEMY

[ICLR 2025]

MolSpectra: Pre-training 3D Molecular Representation with Multi-modal Energy Spectra

Liang Wang^{1,2}, Shaozhen Liu¹, Yu Rong³, Deli Zhao³, Qiang Liu^{1,2}, Shu Wu^{1,2}, Liang Wang^{1,2}

¹Institute of Automation, Chinese Academy of Sciences

²University of Chinese Academy of Sciences

³DAMO Academy, Alibaba Group

工作二：(表征语义增强) 基于分子多模态能量谱感知的分子表征学习

- 分子3D结构去噪等价于学习分子力场 (Denoising as learning a force field)

- It is not feasible to learn molecular force field directly, since it is either unknown or expensive to evaluate.
- Alternative: approximate the data-generating force field with one that can be cheaply evaluated.
- Prove that the denoising objective is equivalent to learning the molecular force field:
 - Molecular structure: $\mathbf{x} \in \mathbb{R}^{3N}$
 - The structure follows the Boltzmann distribution: $p_{\text{physical}}(\mathbf{x}) \propto \exp(-E(\mathbf{x}))$
 - Force field: $\nabla_{\mathbf{x}} \log p_{\text{physical}}(\mathbf{x}) = -\nabla_{\mathbf{x}} E(\mathbf{x})$
 - Approximate p_{physical} with a mixture of Gaussians centered at the known equilibrium structures

$$p_{\text{physical}}(\tilde{\mathbf{x}}) \approx q_{\sigma}(\tilde{\mathbf{x}}) := \frac{1}{n} \sum_{i=1}^n q_{\sigma}(\tilde{\mathbf{x}} | \mathbf{x}_i)$$

where $q_{\sigma}(\tilde{\mathbf{x}} | \mathbf{x}_i) = \mathcal{N}(\tilde{\mathbf{x}}; \mathbf{x}_i, \sigma^2 I_{3N})$.

工作二：(表征语义增强) 基于分子多模态能量谱感知的分子表征学习

- 分子3D结构去噪等价于学习分子力场 (Denoising as learning a force field) (Cont.)

- Learning the force field now yields a score-matching objective:

$$\mathbb{E}_{q_\sigma(\tilde{\mathbf{x}})} [\| \text{GNN}_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}}) \|^2]$$

- According to reference [1], minimizing the following two objectives is equivalent:

$$J_1(\theta) = \mathbb{E}_{q_\sigma(\tilde{\mathbf{x}})} [\| \text{GNN}_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}}) \|^2]$$

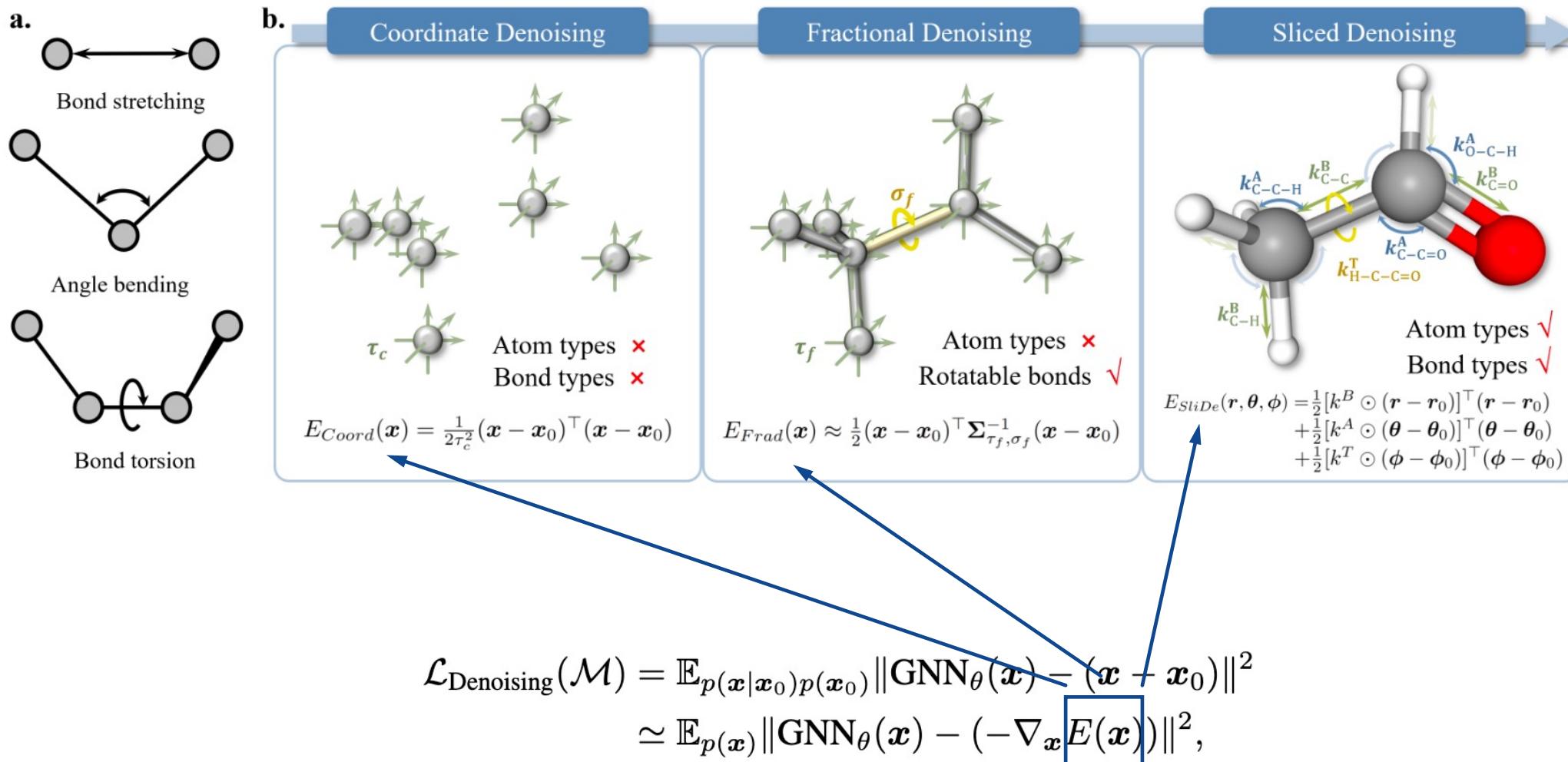
$$J_2(\theta) = \mathbb{E}_{q_\sigma(\tilde{\mathbf{x}}, \mathbf{x})} [\| \text{GNN}_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}} | \mathbf{x}) \|^2]$$

- Thus, the objective in Eq. (1) is equivalent to:

$$\mathbb{E}_{q_\sigma(\tilde{\mathbf{x}}, \mathbf{x})} [\| \text{GNN}_\theta(\tilde{\mathbf{x}}) - \nabla_{\tilde{\mathbf{x}}} \log q_\sigma(\tilde{\mathbf{x}} | \mathbf{x}) \|^2] = \mathbb{E}_{q_\sigma(\tilde{\mathbf{x}}, \mathbf{x})} \left[\| \text{GNN}_\theta(\tilde{\mathbf{x}}) - \frac{\mathbf{x} - \tilde{\mathbf{x}}}{\sigma^2} \|^2 \right]$$

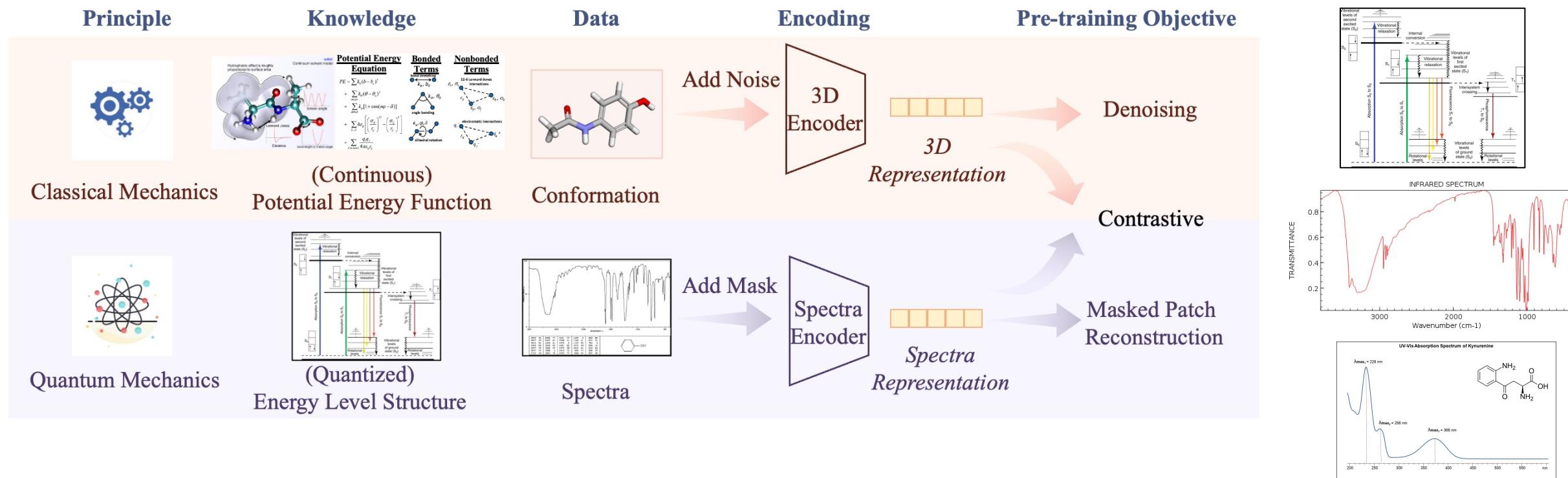
Establishing the relationship between 3D geometries and the energy states of molecular systems is an effective pathway to learn 3D molecular representations.

工作二：(表征语义增强) 基于分子多模态能量谱感知的分子表征学习

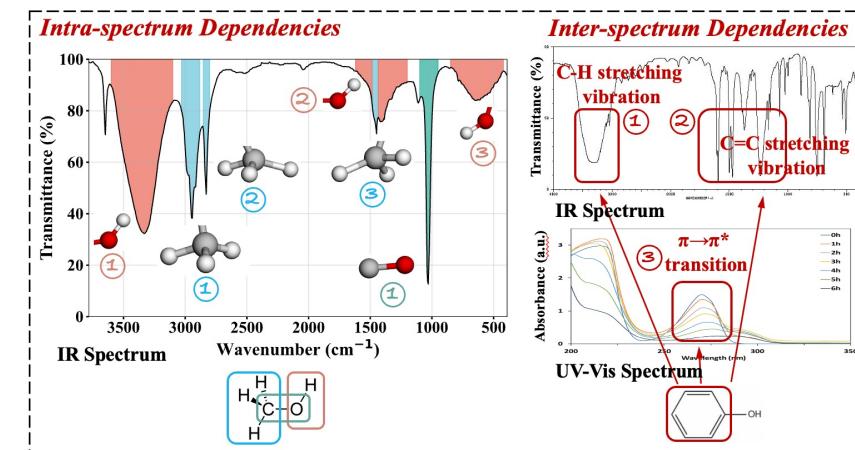
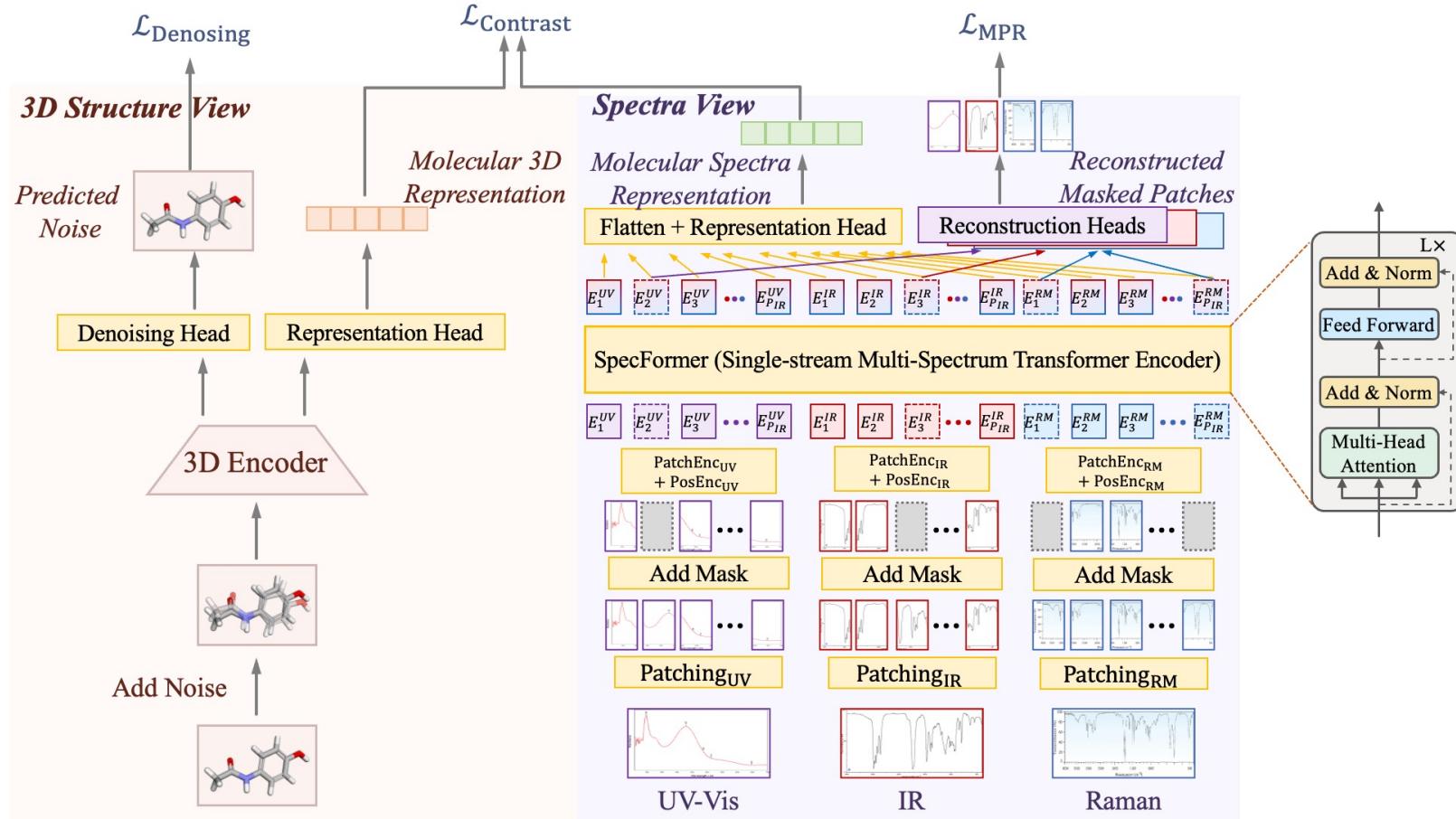


工作二：(表征语义增强) 基于分子多模态能量谱感知的分子表征学习

- 建模分子系统的能量状态是学习分子表征的重要途径。
- 现有方法仅关注经典力学中的连续势能函数，而没有考虑量子力学中的离散能级结构信息。
- 尽管分子能级结构难以获取，但是实验测定的分子能量谱可以反映分子能级跃迁（红外光谱：振动能级跃迁；紫外-可见光谱：电子能级跃迁）。



工作二：(表征语义增强) 基于分子多模态能量谱感知的分子表征学习



光谱内与光谱间依赖关系

$$\mathcal{L} = \beta_{\text{Denoising}} \mathcal{L}_{\text{Denoising}} + \beta_{\text{MPR}} \mathcal{L}_{\text{MPR}} + \beta_{\text{Contrast}} \mathcal{L}_{\text{Contrast}}$$

SpecFormer架构与预训练策略

工作二：(表征语义增强) 基于分子多模态能量谱感知的分子表征学习

Effectiveness of Molecular Spectra in Training from Scratch

Table 1: Performance (MAE \downarrow) when training from scratch on QM9 dataset.

Task Units	μ (D)	α (a_0^3)	homo (meV)	lumo (meV)	gap (meV)	R^2 (a_0^2)	ZPVE (meV)	U_0 (meV)	U (meV)	H (meV)	G (meV)	C_v ($\frac{cal}{mol \cdot K}$)
w/o spectra	0.029	0.071	29	25	48	0.106	1.55	11	12	12	12	0.031
w/ spectra	0.027	0.049	28	24	43	0.084	1.45	10	11	10	10	0.030

工作二：(表征语义增强) 基于分子多模态能量谱感知的分子表征学习

Effectiveness of Molecular Spectra in Representation Pre-Training

Table 2: Performance (MAE \downarrow) on QM9 dataset. The compared methods are divided into two groups training from scratch and pre-training then fine-tuning. The best results are highlighted in bold.

	μ (D)	α (a_0^3)	homo (meV)	lumo (meV)	gap (meV)	R^2 (a_0^2)	ZPVE (meV)	U_0 (meV)	U (meV)	H (meV)	G (meV)	C_v ($\frac{cal}{mol \cdot K}$)
SchNet	0.033	0.235	41.0	34.0	63.0	0.070	1.70	14.00	19.00	14.00	14.00	0.033
EGNN	0.029	0.071	29.0	25.0	48.0	0.106	1.55	11.00	12.00	12.00	12.00	0.031
DimeNet++	0.030	0.044	24.6	19.5	32.6	0.330	1.21	6.32	6.28	6.53	7.56	0.023
PaiNN	0.012	0.045	27.6	20.4	45.7	0.070	1.28	5.85	5.83	5.98	7.35	0.024
SphereNet	0.025	0.045	22.8	18.9	31.1	0.270	1.12	6.26	6.36	6.33	7.78	0.022
TorchMD-Net	0.011	0.059	20.3	17.5	36.1	0.033	1.84	6.15	6.38	6.16	7.62	0.026
Transformer-M	0.037	0.041	17.5	16.2	27.4	0.075	1.18	9.37	9.41	9.39	9.63	0.022
SE(3)-DDM	0.015	0.046	23.5	19.5	40.2	0.122	1.31	6.92	6.99	7.09	7.65	0.024
3D-EMGP	0.020	0.057	21.3	18.2	37.1	0.092	1.38	8.60	8.60	8.70	9.30	0.026
Coord	0.016	0.052	17.7	14.7	31.8	0.450	1.71	6.57	6.11	6.45	6.91	0.020
MolSpectra	0.011	0.048	15.5	13.1	26.8	0.410	1.71	5.67	5.45	5.87	6.18	0.021

工作二：(表征语义增强) 基于分子多模态能量谱感知的分子表征学习

Effectiveness of Molecular Spectra in Representation Pre-Training

Table 3: Performance (MAE \downarrow) on MD17 force prediction (kcal/mol/ Å). The methods are divided into two groups: training from scratch and pre-training then fine-tuning. The best results are in bold.

	Aspirin	Benzene	Ethanol	Malonal -dehyde	Naphtha -lene	Salicy -lic Acid	Toluene	Uracil
SphereNet	0.430	0.178	0.208	0.340	0.178	0.360	0.155	0.267
SchNet	1.350	0.310	0.390	0.660	0.580	0.850	0.570	0.560
DimeNet	0.499	0.187	0.230	0.383	0.215	0.374	0.216	0.301
PaiNN	0.338	-	0.224	0.319	0.077	0.195	0.094	0.139
TorchMD-Net	0.245	0.219	0.107	0.167	0.059	0.128	0.064	0.089
SE(3)-DDM*	0.453	-	0.166	0.288	0.129	0.266	0.122	0.183
Coord	0.211	0.169	0.096	0.139	0.053	0.109	0.058	0.074
MolSpectra	0.099	0.097	0.052	0.077	0.085	0.093	0.075	0.095

工作二：(表征语义增强) 基于分子多模态能量谱感知的分子表征学习

Sensitivity Analysis of Patch Length, Stride, and Mask Ratio

Table 4: Sensitivity of patch length and stride.

patch length	stride	overlap ratio	homo	lumo	gap
20	5	75%	15.9	13.7	28.0
20	10	50%	15.5	13.1	26.8
20	15	25%	16.1	13.6	28.1
20	20	0%	15.7	13.5	27.5
16	8	50%	16.0	13.4	27.6
30	15	50%	15.9	14.0	28.1

Table 5: Sensitivity of mask ratio.

mask ratio	homo	lumo	gap
0.05	15.7	13.4	29.7
0.10	15.5	13.1	26.8
0.15	15.7	13.5	28.0
0.20	16.0	13.6	28.1
0.25	16.3	13.5	28.0
0.30	16.2	13.7	29.0

Ablation Study of Spectral Modalities

Table 7: Ablation of spectral modalities.

UV-Vis	IR	Raman	homo	lumo	gap
✓	✓	✓	15.5	13.1	26.8
-	✓	✓	15.8	13.3	27.1
✓	-	✓	16.6	14.1	28.9
✓	✓	-	16.1	13.9	28.3

工作二：(表征语义增强) 基于分子多模态能量谱感知的分子表征学习

Visualization of Attention Patterns and Learned Spectra Representations in SpecFormer

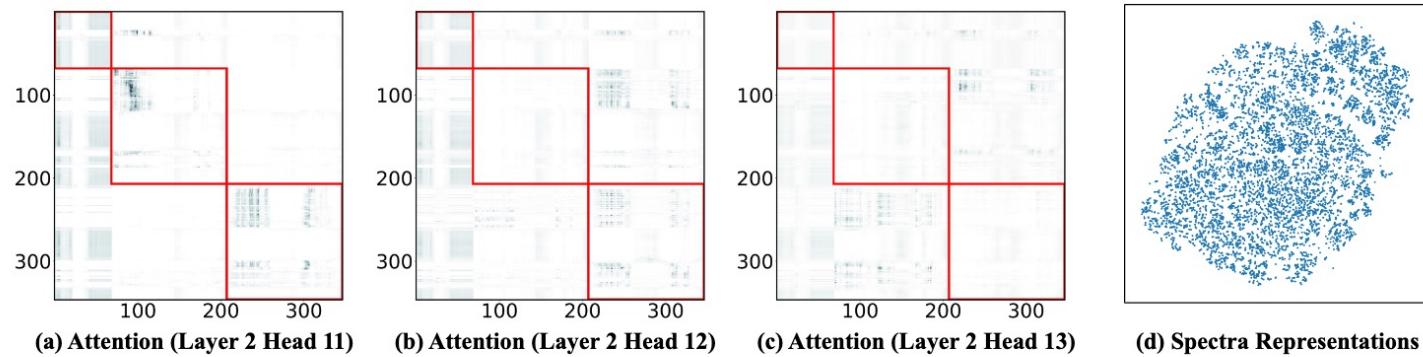


Figure A2: (a-c) Attention maps from three attention heads in SpecFormer. Different heads model distinct dependencies. (d) t-SNE visualization of the spectra representations produced by SpecFormer.

目录

1 / 研究背景

2 / 理解：分子表征学习

3 / 预测：分子性质预测

4 / 生成：分子结构解析



中国科学院自动化研究所
模式识别实验室
New Laboratory of Pattern Recognition



多模态人工智能系统
全国重点实验室
State Key Laboratory of
Multimodal Artificial Intelligence Systems



中国科学院自动化研究所
Institute of Automation
Chinese Academy of Sciences



NUS
National University
of Singapore

[NeurIPS 2024]
**Pin-Tuning: Parameter-Efficient In-Context Tuning
for Few-Shot Molecular Property Prediction**

Liang Wang^{1,2}, Qiang Liu^{1,2}, Shaozhen Liu³, Xin Sun⁴, Shu Wu^{1,2}, Liang Wang^{1,2,4}

¹Institute of Automation, Chinese Academy of Sciences

²University of Chinese Academy of Sciences

³Beijing Institute of Technology

⁴University of Science and Technology of China

工作三：基于参数高效和上下文感知微调的小样本分子性质预测

Key Elements underlying Molecular Property Prediction

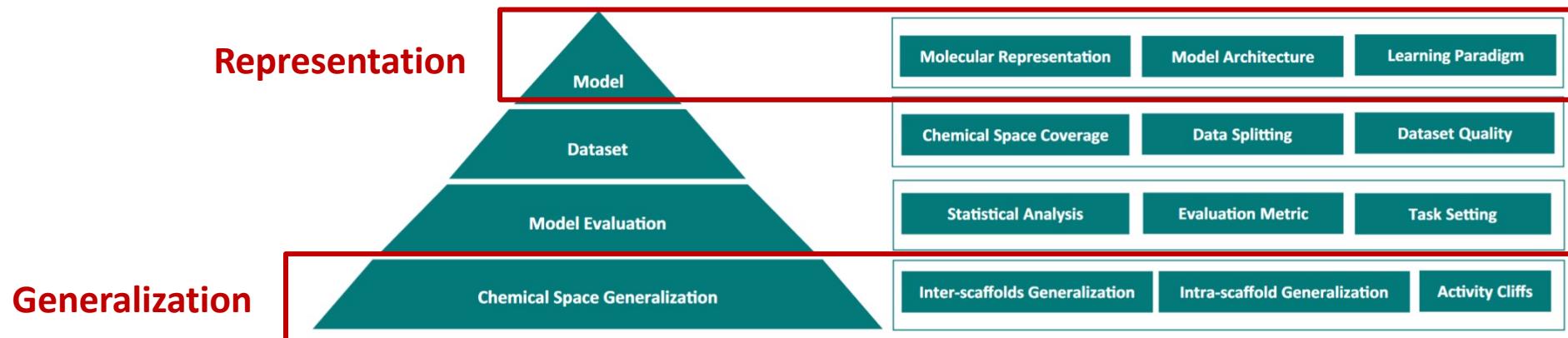
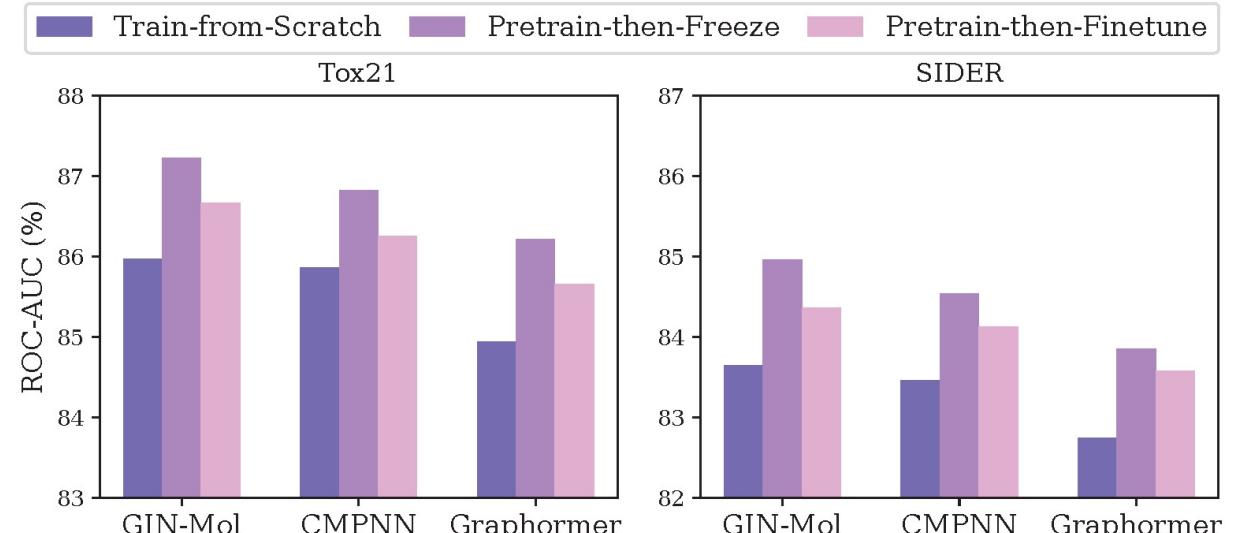
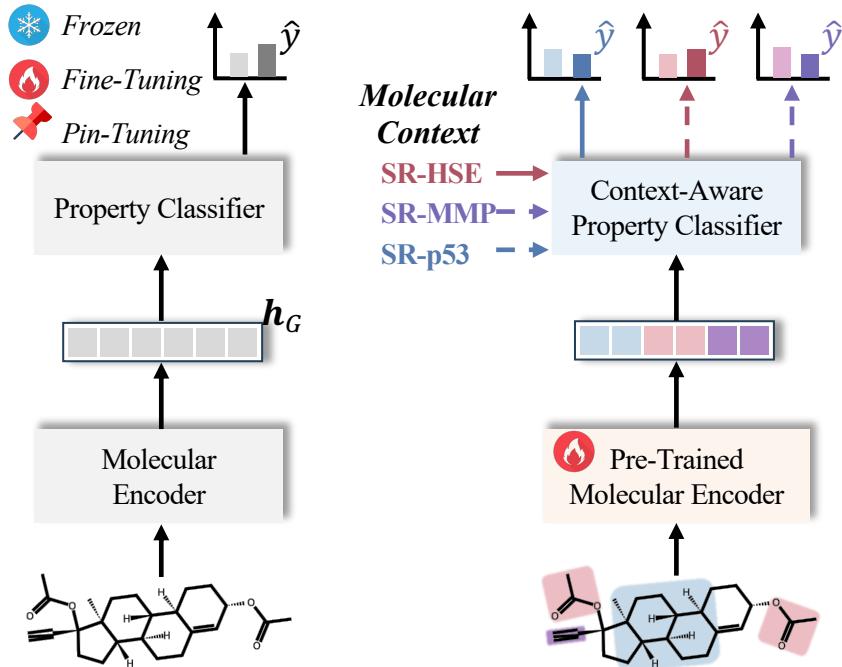


Fig. 1 | Key elements underlying molecular property prediction. There are four aspects involved: model, dataset, model evaluation, and generalization. In the literature, the focus is more on the model, which aims at developing novel learning paradigms or model architectures on certain molecular representations. However, it is also necessary to consider other crucial elements, pertaining to (1) what the model is built upon, (2) how the model is evaluated, and (3) eventually what the model is capable of. For the dataset, its chemical space coverage (w.r.t. both structures and

labels), and scrutiny of its quality, including dataset size and label accuracy (e.g., duplicates, contradictions, and noise), as well as data splitting, is essential before developing a model for a specific property prediction task. For the model evaluation, thoughtful consideration of statistical analysis, evaluation metrics, and task settings is critical as they impact the observed prediction performance. For the chemical space generalization, it is important to clarify the model's applicability and if the activity-cliffs issue is addressed.

工作三：基于参数高效和上下文感知微调的小样本分子性质预测

Few-Shot Molecular Property Prediction



Observation

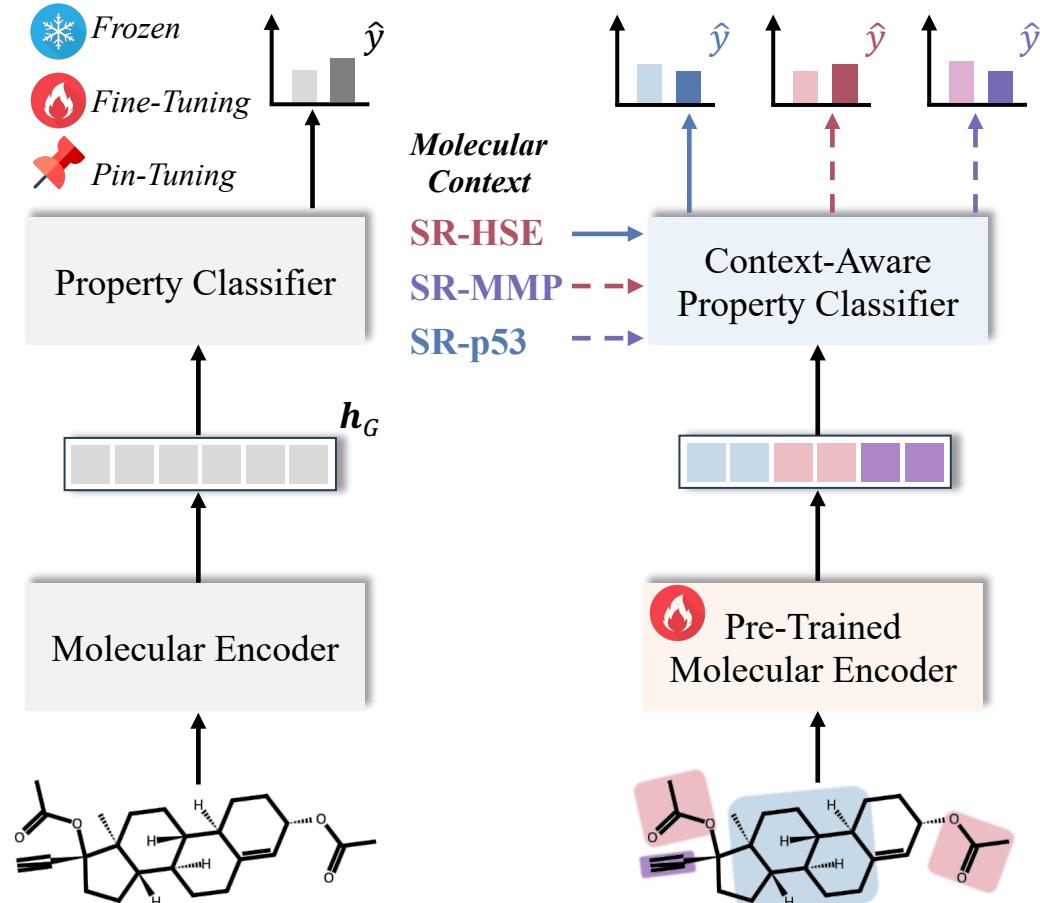
Train-from-Scratch < Pretrain-then-Finetune \leq Pretrain-then-Freeze

Pre-training is effective, but fine-tuning is ineffective.

*How to adapt molecular pre-trained models to downstream tasks,
especially in few-shot scenarios?*

工作三：基于参数高效和上下文感知微调的小样本分子性质预测

Few-Shot Molecular Property Prediction



已有方法无法有效微调的原因：

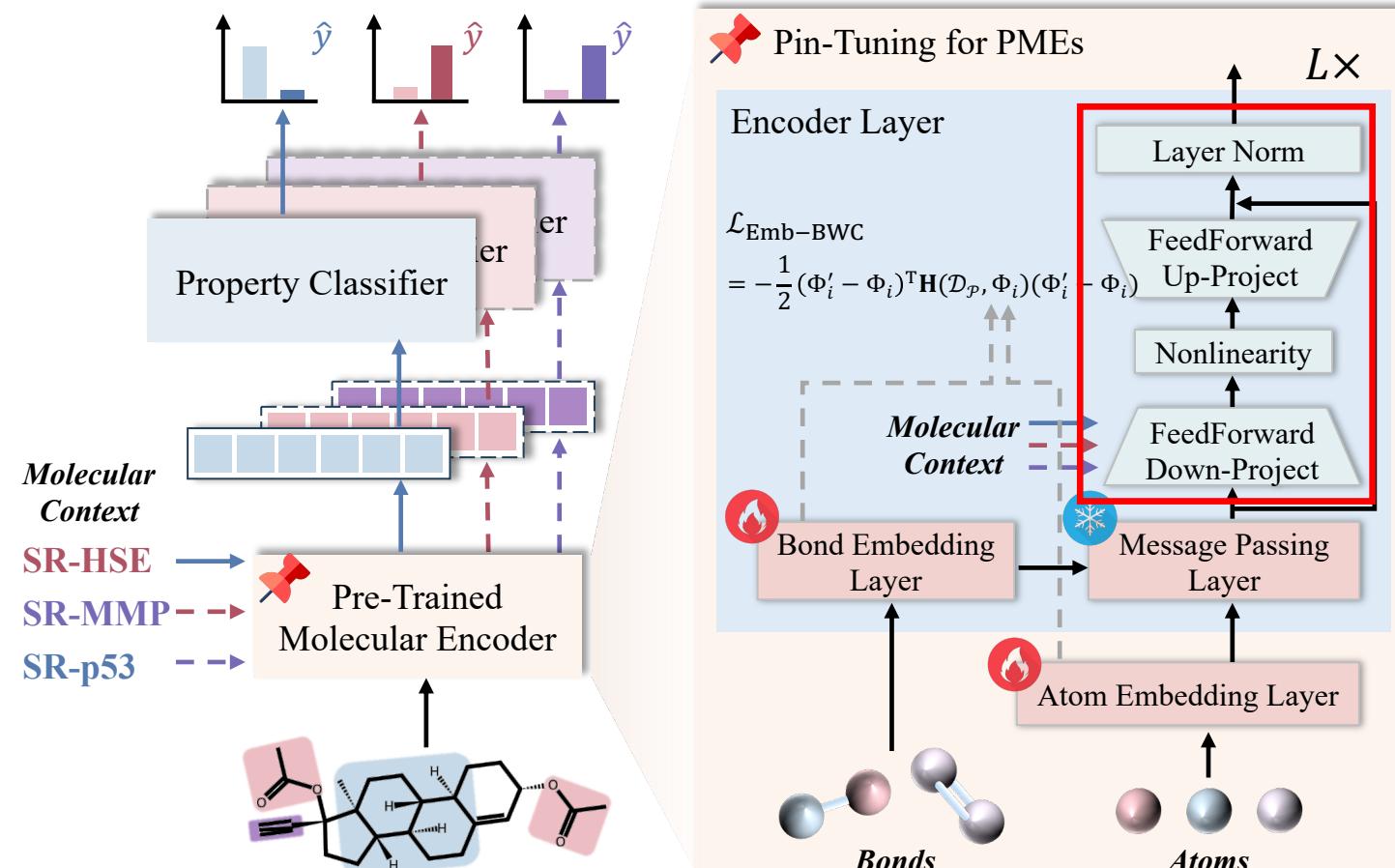
1. 数据量与参数量之间不平衡
2. 分子上下文感知能力有限

(a) Vanilla MPP framework.

(b) Existing FSMPP framework.

工作三：基于参数高效和上下文感知微调的小样本分子性质预测

Pin-Tuning: Parameter-Efficient In-Context Tuning for Few-Shot Molecular Property Prediction



MP-Adapter: message passing layer-oriented adapter

$$z_v^{(l)} = \text{FeedForward}_{\text{down}}(\mathbf{h}_v^{(l)}) \in \mathbb{R}^{d_2},$$

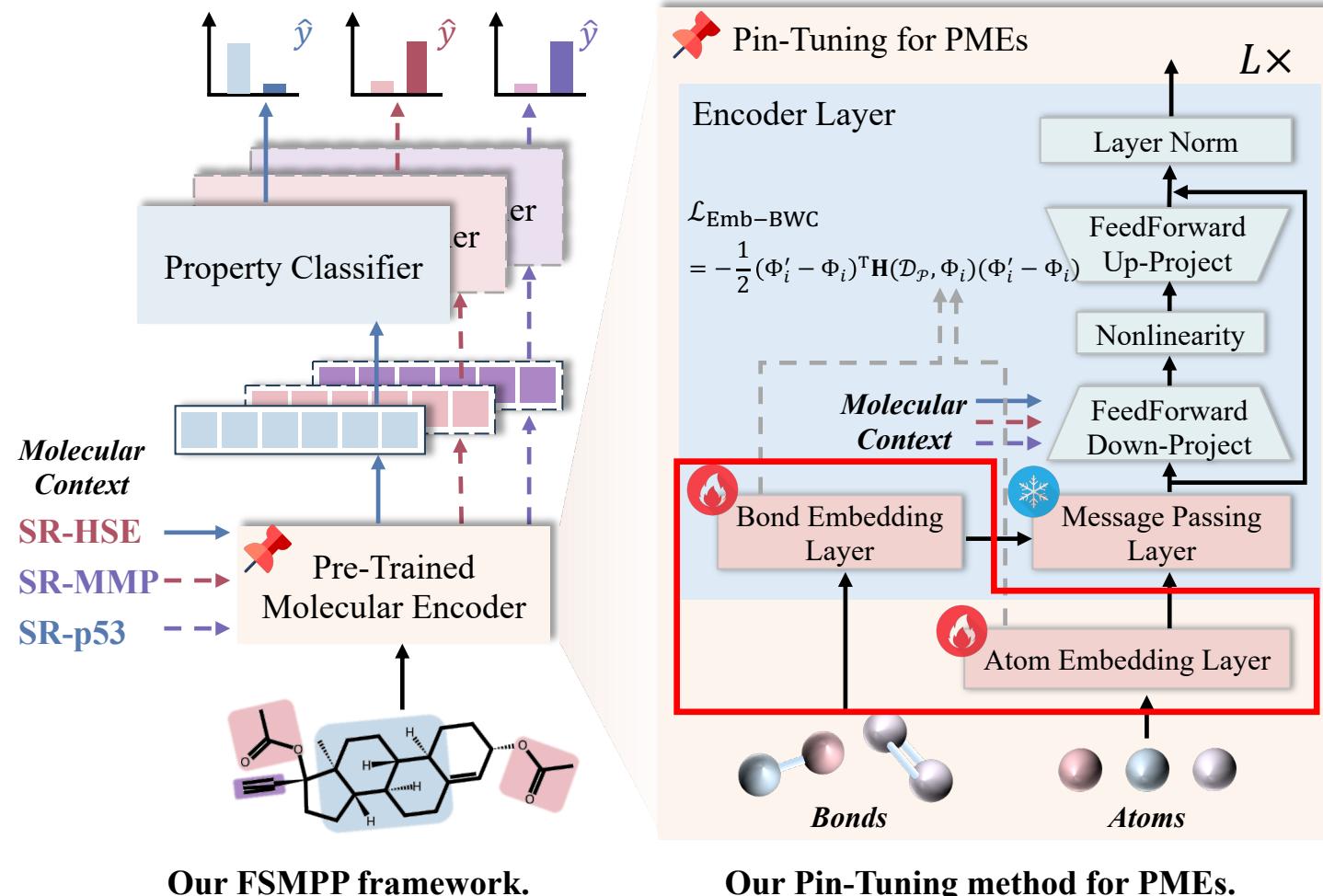
$$\Delta \mathbf{h}_v^{(l)} = \text{FeedForward}_{\text{up}}(\phi(z_v^{(l)})) \in \mathbb{R}^d,$$

$$\tilde{\mathbf{h}}_v^{(l)} = \text{LayerNorm}(\mathbf{h}_v^{(l)} + \Delta \mathbf{h}_v^{(l)}) \in \mathbb{R}^d,$$

- *Bottleneck*
- *Near-zero initialization*
- *Skip-connection*

工作三：基于参数高效和上下文感知微调的小样本分子性质预测

Pin-Tuning: Parameter-Efficient In-Context Tuning for Few-Shot Molecular Property Prediction



Emb-BWC: embedding layer-oriented Bayesian weight consolidation

$$\mathcal{L}_{\text{Emb-BWC}} = -\frac{1}{2} \sum_{i=1}^E (\Phi'_i - \Phi_i)^\top \mathbf{H}(\mathcal{D}_P, \Phi_i)(\Phi'_i - \Phi_i),$$

- Maximum a posterior (MAP) estimation
- Bayesian learning theory
- Second-order Taylor expansion

Three choices of diagonal approximation of Hessian

$$\mathcal{L}_{\text{Emb-BWC}}^{\text{IM}} = \frac{1}{2} \sum_{i=1}^E \sum_{j=1}^d (\Phi'_{i,j} - \Phi_{i,j})^2$$

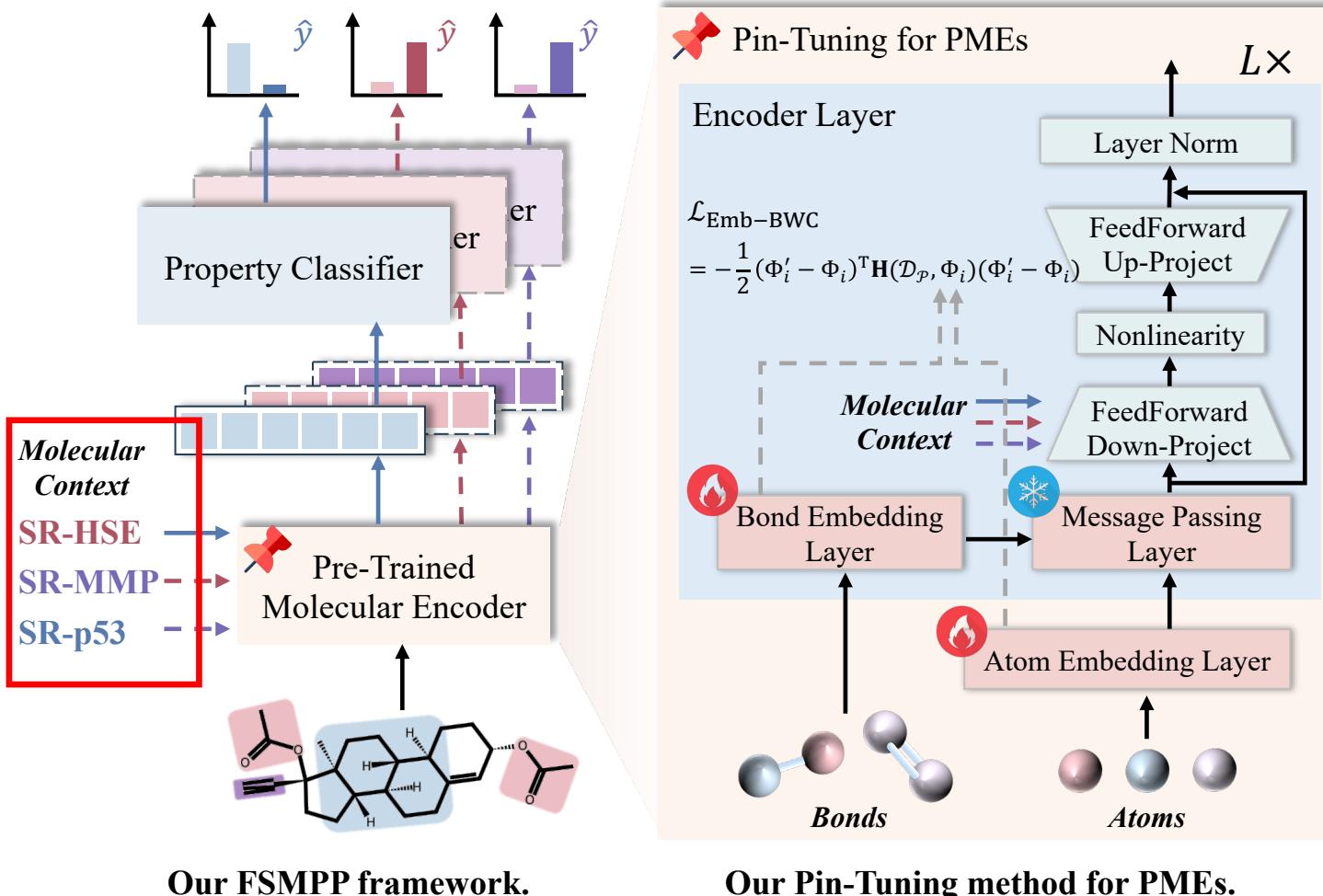
$$\mathcal{L}_{\text{Emb-BWC}}^{\text{FIM}} = \frac{1}{2} \sum_{i=1}^E \tilde{\mathbf{F}}_i (\Phi'_i - \Phi_i)^2$$

$$\mathcal{L}_{\text{Emb-BWC}}^{\text{EFIM}} = \frac{1}{2} \sum_{i=1}^E \tilde{\mathbf{F}}_i (\tilde{\Phi}'_i - \tilde{\Phi}_i)^2$$

- Identity matrix.
- Diagonal of Fisher information matrix.
- Diagonal of embedding-wise Fisher information matrix.

工作三：基于参数高效和上下文感知微调的小样本分子性质预测

Pin-Tuning: Parameter-Efficient In-Context Tuning for Few-Shot Molecular Property Prediction



Enabling contextual perceptiveness in MP-Adapter

	p_t	p_1^{seen}	p_2^{seen}	p_3^{seen}
m_1^s	1	1	0	?
m_2^s	1	0	0	1
m_3^s	0	0	?	0
m_4^s	0	1	1	0
m_1^q	?	0	0	1
m_2^q	?	1	1	0

Context Graph

Legend: Context Information (light blue), Molecule (light orange), Property (pink), Positive Label (solid line), Negative Label (dashed line), Unknown Label (dotted line).

Figure 3: Convert the context information of a 2-shot episode into a context graph.

$$\mathbf{C} = \text{ContextEncoder}(\mathcal{V}_t, \mathbf{A}_t, \mathbf{X}_t)$$

$$z^{(l)} = \text{FeedForward}_{\text{down}}(\mathbf{h}_v^{(l)} \| \mathbf{c}_m \| \mathbf{c}_p),$$

工作三：基于参数高效和上下文感知微调的小样本分子性质预测

Experiment Results

Table 1: ROC-AUC scores (%) on benchmark datasets, compared with methods trained from scratch (first group) and methods that leverage pre-trained molecular encoder (second group). The best is marked with **boldface** and the second best is with underline. $\Delta Improve.$ indicates the relative improvements over the baseline models in percentage.

Model	Tox21		SIDER		MUV		ToxCast		PCBA	
	10-shot	5-shot								
Siamese	80.40(0.35)	-	71.10(4.32)	-	59.96(5.13)	-	-	-	-	-
ProtoNet	74.98(0.32)	72.78(3.93)	64.54(0.89)	64.09(2.37)	65.88(4.11)	64.86(2.31)	68.87(0.43)	66.26(1.49)	64.93(1.94)	62.29(2.12)
MAML	80.21(0.24)	69.17(1.34)	70.43(0.76)	60.92(0.65)	63.90(2.28)	63.00(0.61)	68.30(0.59)	67.56(1.53)	66.22(1.31)	65.25(0.75)
TPN	76.05(0.24)	75.45(0.95)	67.84(0.95)	66.52(1.28)	65.22(5.82)	65.13(0.23)	69.47(0.71)	66.04(1.14)	67.61(0.33)	63.66(1.64)
EGNN	81.21(0.16)	76.80(2.62)	72.87(0.73)	60.61(1.06)	65.20(2.08)	63.46(2.58)	74.02(1.11)	67.13(0.50)	69.92(1.85)	67.71(3.67)
IterRefLSTM	81.10(0.17)	-	69.63(0.31)	-	49.56(5.12)	-	-	-	-	-
Pre-GNN	82.14(0.08)	82.04(0.30)	73.96(0.08)	76.76(0.53)	67.14(1.58)	70.23(1.40)	75.31(0.95)	74.43(0.47)	76.79(0.45)	75.27(0.49)
Meta-MGNN	82.97(0.10)	76.12(0.23)	75.43(0.21)	66.60(0.38)	68.99(1.84)	64.07(0.56)	76.27(0.56)	75.26(0.43)	72.58(0.34)	72.51(0.52)
PAR	84.93(0.11)	83.95(0.15)	78.08(0.16)	77.70(0.34)	<u>69.96</u> (1.37)	<u>68.08</u> (2.42)	79.41(0.08)	76.89(0.32)	73.71(0.61)	72.79(0.98)
GS-Meta	<u>86.67</u> (0.41)	<u>86.43</u> (0.02)	<u>84.36</u> (0.54)	<u>84.57</u> (0.01)	66.08(1.25)	64.50(0.20)	<u>83.81</u> (0.16)	<u>82.65</u> (0.35)	<u>79.40</u> (0.43)	<u>77.47</u> (0.29)
Pin-Tuning	91.56 (2.57)	90.95 (2.33)	93.41 (3.52)	92.02 (3.01)	73.33 (2.00)	70.71 (1.42)	84.94 (1.09)	83.71 (0.93)	81.26 (0.46)	79.23 (0.52)
$\Delta Improve.$	5.64%	5.23%	10.73%	8.81%	4.82%	3.86%	1.35%	1.28%	2.34%	2.27%

工作三：基于参数高效和上下文感知微调的小样本分子性质预测

微调参数量分析

$$N_{Fine-Tuning} = |E_n|d + L(|E_e|d + 2dd_1 + 3d + d_1).$$

$$N_{Pin-Tuning} = |E_n|d + L(|E_e|d + 2dd_2 + 3d + d_2).$$

$$\Delta N = (d_1 - d_2)L(2d + 1).$$

Ours (14.2% parameters, higher performance)

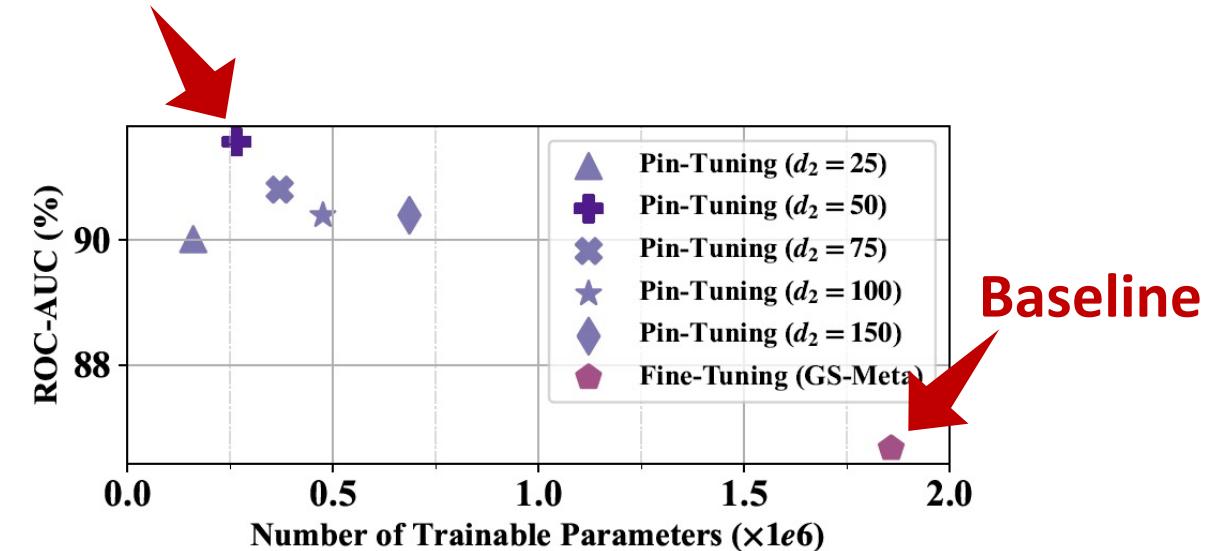
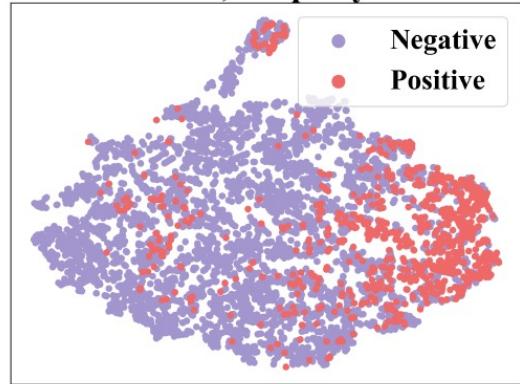


Figure 5: ROC-AUC (%) and number of trainable parameters of Pin-Tuning with varied value of d_2 and full Fine-Tuning method (e.g., GS-Meta) on the Tox21 dataset.

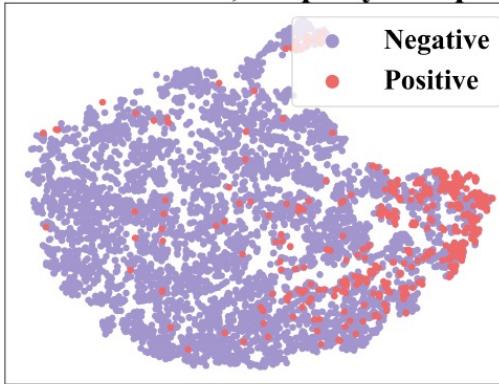
工作三：基于参数高效和上下文感知微调的小样本分子性质预测

表征可视化

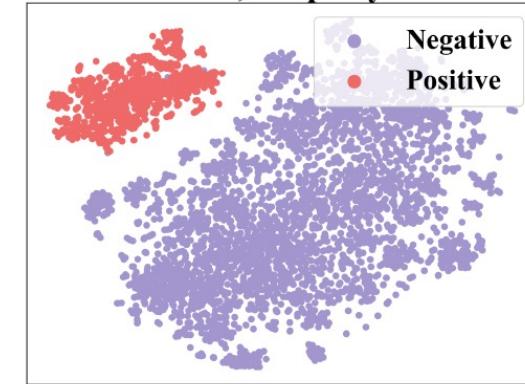
Dataset: Tox21, Property: SR-MMP



Dataset: Tox21, Property: SR-p53



Dataset: Tox21, Property: SR-MMP



Dataset: Tox21, Property: SR-p53

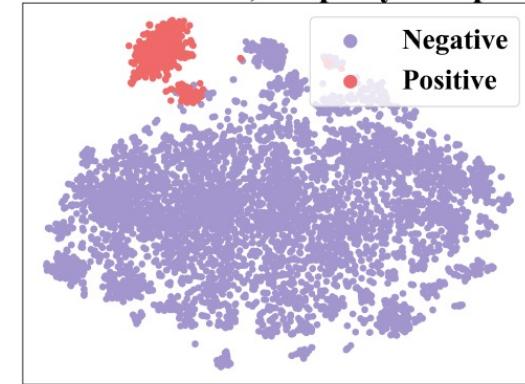


Figure 6: Molecular representations encoded by GS-Meta [58].

Full Fine-Tuning

Figure 7: Molecular representations encoded by Pin-Tuning.

Pin-Tuning

目录

1 / 研究背景

2 / 理解：分子表征学习

3 / 预测：分子性质预测

4 / 生成：分子结构解析



中国科学院自动化研究所
模式识别实验室
New Laboratory of Pattern Recognition



多模态人工智能系统
全国重点实验室
State Key Laboratory of
Multimodal Artificial Intelligence Systems



中国科学院自动化研究所
Institute of Automation
Chinese Academy of Sciences



NUS
National University
of Singapore



[arXiv 2025] DiffSpectra: Molecular Structure Elucidation from Spectra using Diffusion Models

Liang Wang^{1,2,3}, Yu Rong^{4,5}, Tingyang Xu^{4,5}, Zhenyi Zhong⁶, Zhiyuan Liu³,
Pengju Wang^{4,5}, Deli Zhao^{4,5}, Qiang Liu^{1,2}, Shu Wu^{1,2}, Liang Wang^{1,2}

¹Institute of Automation, Chinese Academy of Sciences

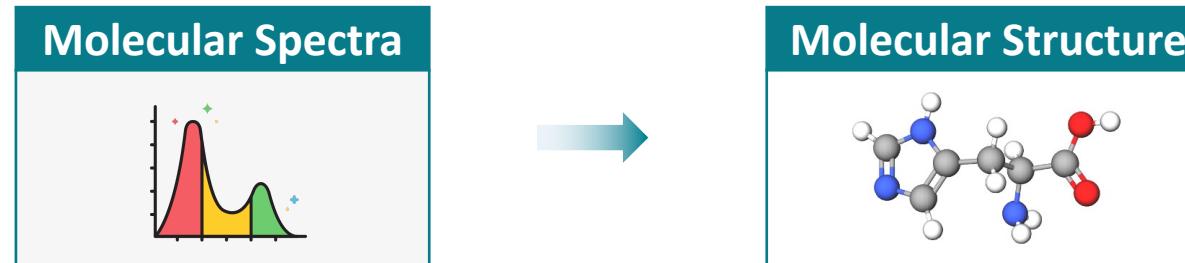
²University of Chinese Academy of Sciences

³National University of Singapore

⁴DAMO Academy, Alibaba Group ⁵Hupan Lab ⁶Tianjin University

Background

Molecular Structure Elucidation (/ Determination / Identification)

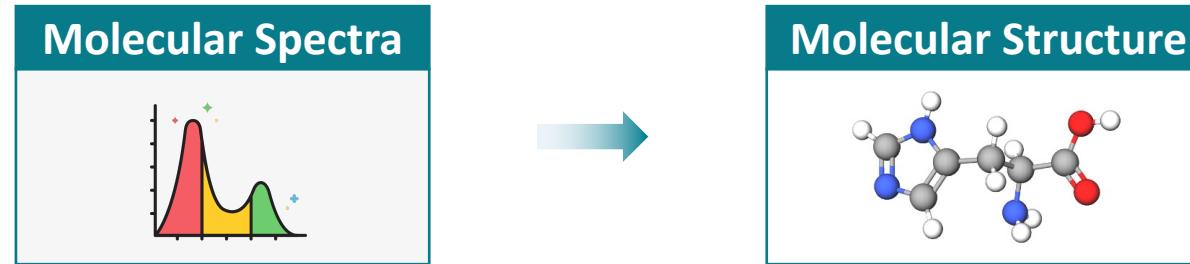


1. What is Molecular Structure Elucidation?

- **Problem Definition:** Molecular structure elucidation is the process of **determining the complete structure of a molecule based on indirect experimental observations**, such as spectroscopy (e.g., IR, UV), mass spectrometry, diffraction (e.g., X-ray diffraction), and cryo-electron microscopy (cryo-EM).
- **Input of the model:** Molecular Spectra (or other kinds of experimental observations)
- **Output of the model:** Molecular Structure

Background

Molecular Structure Elucidation



2. Where is Molecular Structure Elucidation Used?

- When the molecular structure of a substance is unknown—such as a product of a chemical reaction or a compound collected from the natural environment—we rely on structure elucidation techniques to identify its full or partial structural information.

3. This is a universal problem across disciplines:

Research Field

Molecule Types



Chemistry



Small Molecules



Biology



Proteins



Materials



Crystals

Structure Elucidation Techniques

IR (Infrared Spectra, 红外光谱), UV-Vis (Ultraviolet-Visible Spectra, 紫外光谱),

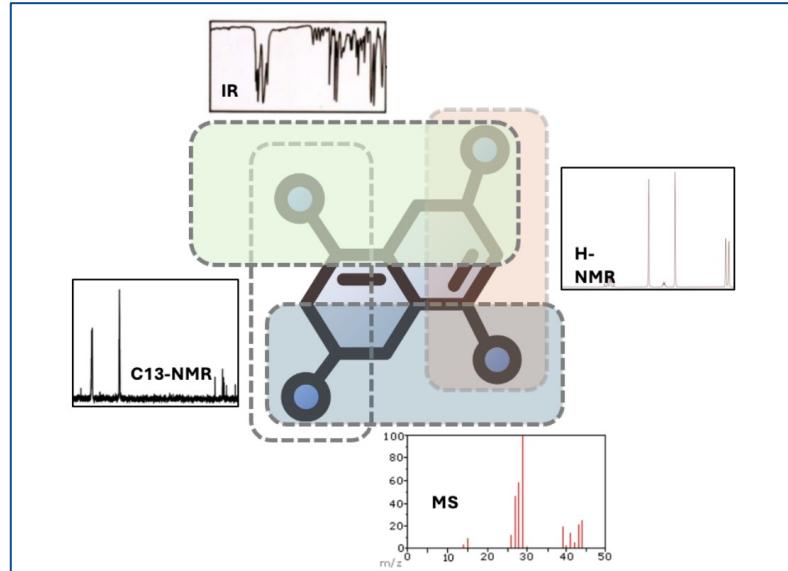
Raman (Raman Spectra, 拉曼光谱), MS (Mass Spectra, 质谱), NMR (Nuclear Magnetic Resonance, 核磁共振谱)

MS (Mass Spectra, 质谱), XRD (X-Ray Diffraction, X射线衍射), Cryo-EM (Cryogenic Electron Microscopy, 冷冻电镜)

XRD (X-Ray Diffraction, X射线衍射)

Background

Molecular Structure Elucidation



4. Different types of spectra are complementary, capturing different structural aspects

Spectrum Type	Encodes Information About
IR	Vibrational frequencies, functional groups
Raman	Molecular symmetry, bond strength, shape
UV-Vis	electronic transitions

Single spectrum is often insufficient

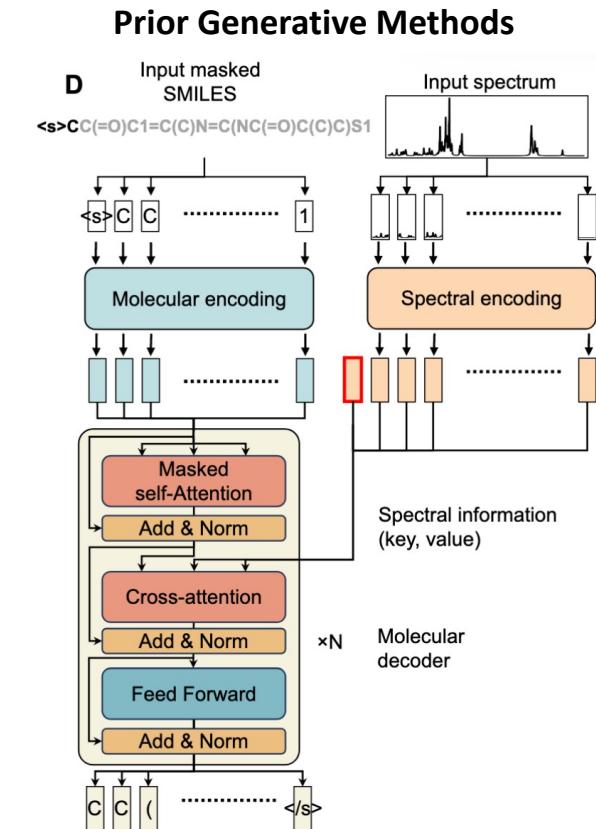
- Individual spectra give **partial clues**
 - Real-world structure elucidation **requires integrating multiple spectra**
- Multi-spectral fusion is essential for accurate structure elucidation**

Motivation

Molecular Structure Elucidation

Existing Methods and Their Limitations

- Traditional Methods: Human Expert-Driven
 - Manual peak assignment, hypothesis testing, and iteration
 - ✗ Time-consuming, labor-intensive, subjective, non-scalable
- Early Machine Learning Methods: Retrieval-Based
 - Match input spectra to a candidate structure in an established reference library
 - Their performance **relies on finite reference library**
 - ✓ automatic and efficient
 - ✗ Rely on the coverage of established library, cannot generalize to out-of-distribution molecules
- Prior Generative Methods
 - Autoregressive models
 - Output in SMILES format only
 - ✓ Doesn't rely on finite library, can generalize better
 - ✗ Limited to sequential token generation, not holistic structure
 - ✗ No topological and geometric inductive bias
 - ✗ Can only use a single spectral modality

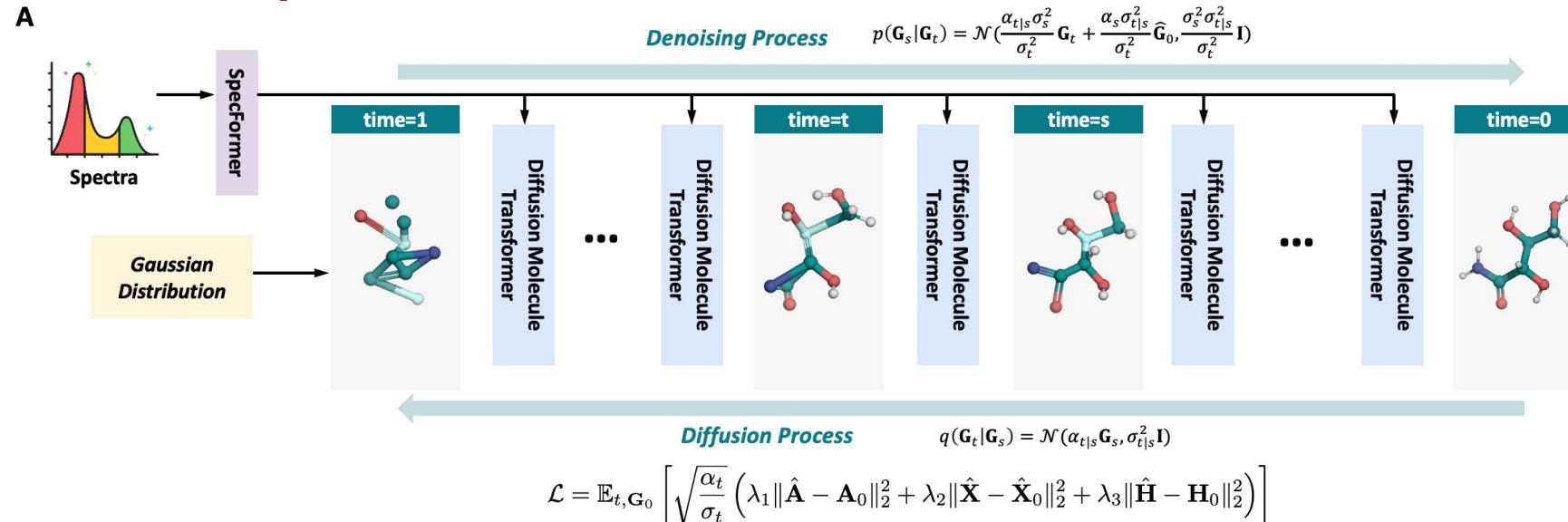


Our DiffSpectra

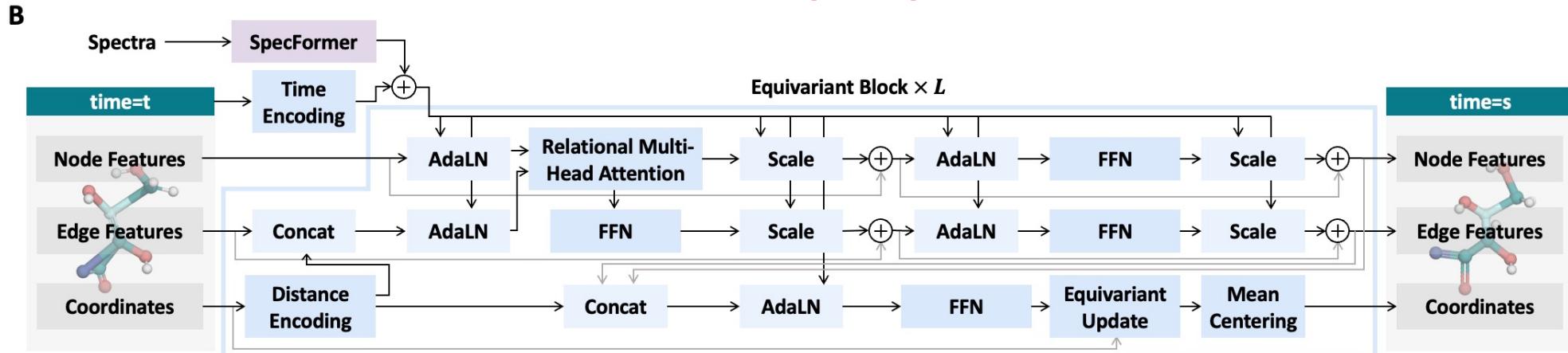
- > diffusion models
- > SE(3)-equivariant denoising network
- > multi-modal spectra encoder

Method: DiffSpectra

Overview of DiffSpectra Framework

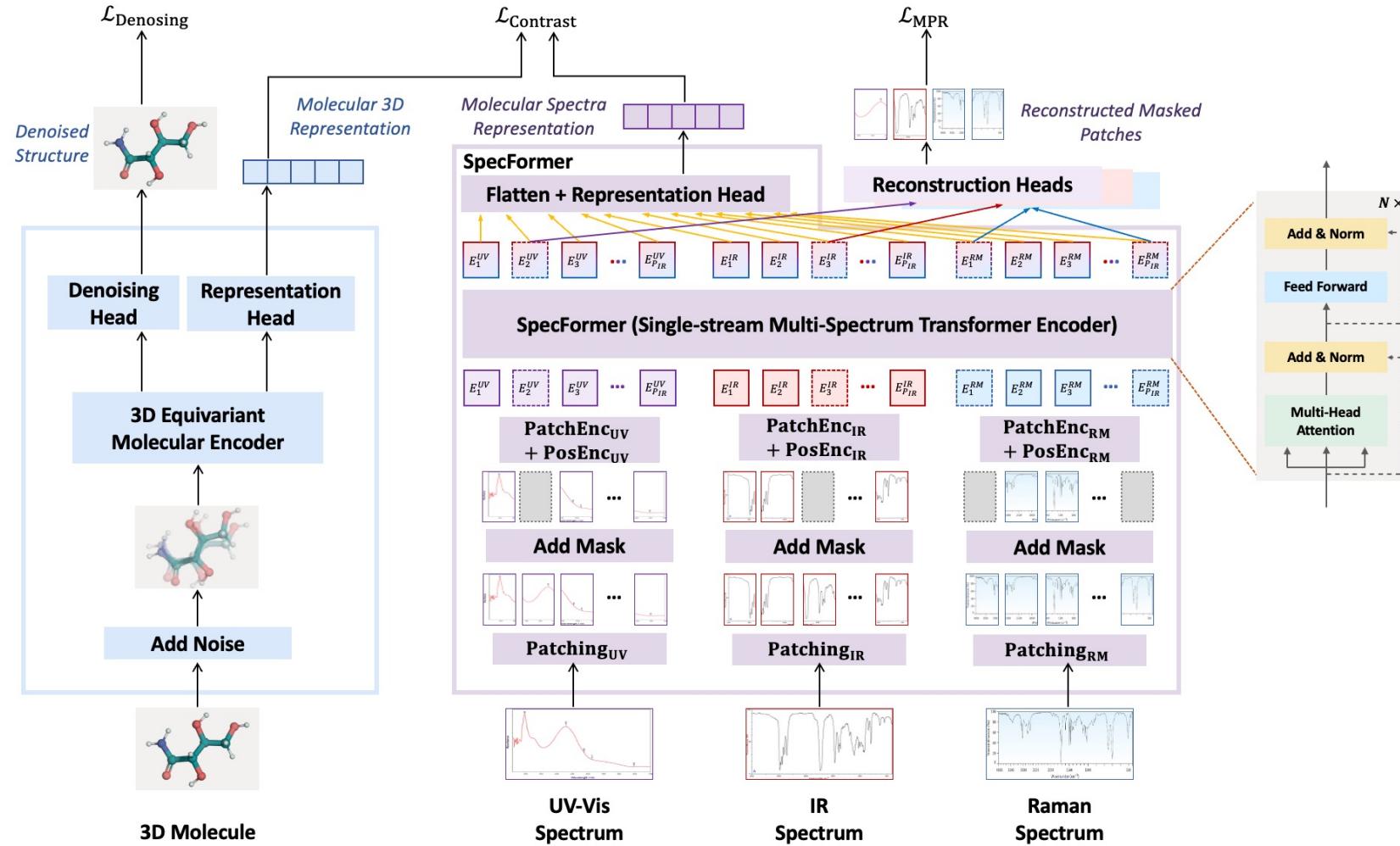


Architecture of Diffusion Molecule Transformer (DMT)



Method: DiffSpectra

Architecture and Pre-Training Strategy of SpecFormer

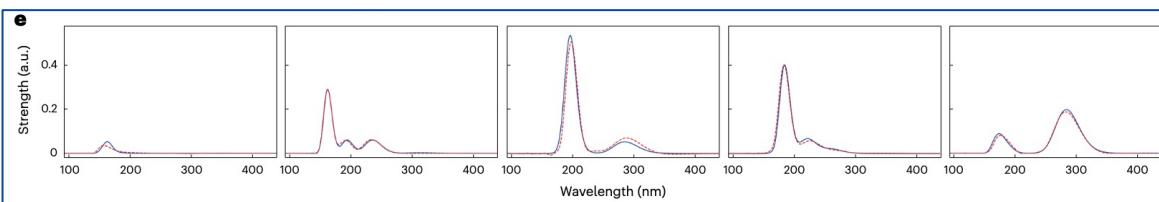
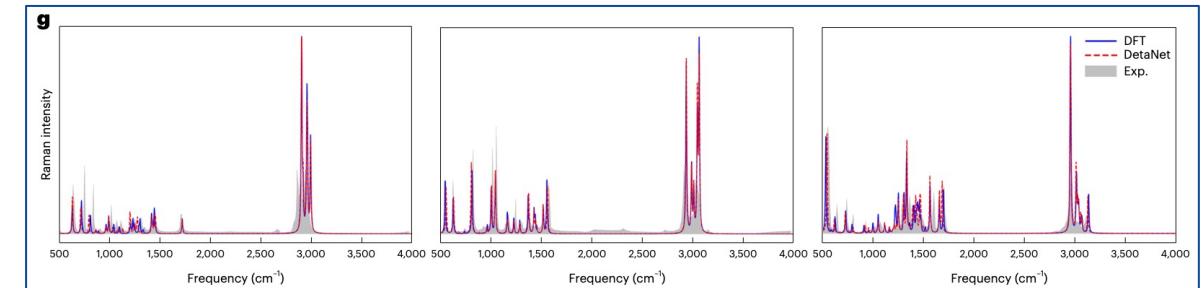
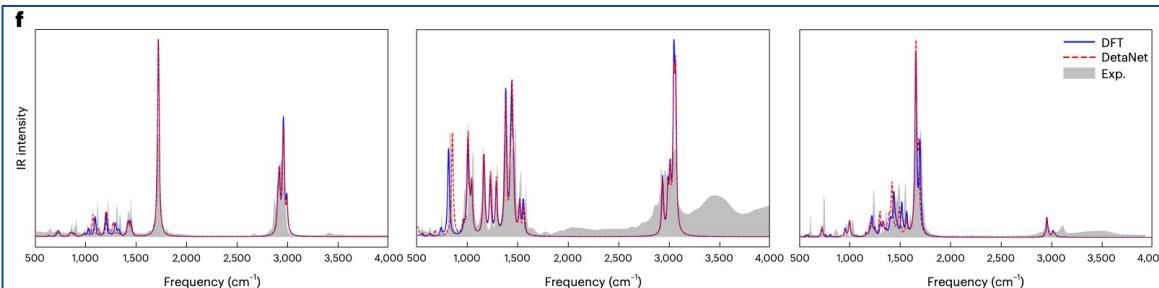


Experimental Results

Dataset

QM9Spectra (QM9S)

- It is proposed in [1] *A deep learning model for predicting selected organic molecular spectra*, *Nature Computational Science*, Nov 2023.
- It is derived from over **130,000 organic molecules** in the original QM9 database.
- Each molecule in the QM9S dataset was subjected to DFT (B3LYP/def-TZVP) calculations to obtain its **IR, Raman, and UV-Vis** spectra. The dataset includes **high-quality molecular geometries** and corresponding spectral information for all three types.



Experimental Results

Metrics

The first group of metrics: used to evaluate performance from the perspective of de novo molecular generation.

Molecular Structure Validity and Stability Evaluation

- Validity and Complete (V&C), Validity and Unique (V&U), Validity, Unique, and Novelty (V&U&N)
- Atom Stability, Mol Stability

Distribution-based Structure Evaluation

- Frechet ChemNet Distance (FCD)
- Similarity to Nearest Neighbor (SNN)
- Fragment Similarity (Frag)
- Scaffold Similarity (Scaf)

3D Geometry-based Structural Evaluation

- Bond Maximum Mean Discrepancy (MMD)
- Angle Maximum Mean Discrepancy (MMD)
- Dihedral Maximum Mean Discrepancy (MMD)

Experimental Results

Metrics

The second group of metrics: used to evaluate performance from the perspective of spectra-conditional molecular structure elucidation.

Top-K Accuracy

$$\text{ACC}@K = \mathbb{E}_{\mathcal{G}} \left[\mathbb{1} \left(\exists i \in \{1, \dots, K\}, \hat{\mathcal{G}}_i = \mathcal{G} \right) \right]$$

Maximum Common Edge Subgraph (MCES)

$$\text{MCES} = \mathbb{E}_{\mathcal{G}} \left[|E(\mathcal{G})| + |E(\hat{\mathcal{G}})| - 2 \cdot \max_{\mathcal{H} \subseteq \mathcal{G}, \mathcal{H} \subseteq \hat{\mathcal{G}}} |E(\mathcal{H})| \right]$$

Fingerprint-based Similarity Metrics

$$\text{TaniSim}_{\text{Morgan}} = \mathbb{E}_{\mathcal{G}} \left[\frac{|\mathbf{a}_{\text{Morgan}} \wedge \hat{\mathbf{a}}_{\text{Morgan}}|}{|\mathbf{a}_{\text{Morgan}} \vee \hat{\mathbf{a}}_{\text{Morgan}}|} \right]$$

$$\text{CosSim}_{\text{Morgan}} = \mathbb{E}_{\mathcal{G}} \left[\frac{\mathbf{a}_{\text{Morgan}} \cdot \hat{\mathbf{a}}_{\text{Morgan}}}{\|\mathbf{a}_{\text{Morgan}}\| \|\hat{\mathbf{a}}_{\text{Morgan}}\|} \right]$$

$$\text{TaniSim}_{\text{MACCS}} = \mathbb{E}_{\mathcal{G}} \left[\frac{|\mathbf{a}_{\text{MACCS}} \wedge \hat{\mathbf{a}}_{\text{MACCS}}|}{|\mathbf{a}_{\text{MACCS}} \vee \hat{\mathbf{a}}_{\text{MACCS}}|} \right]$$

Fragment-based Similarity (Fraggle)

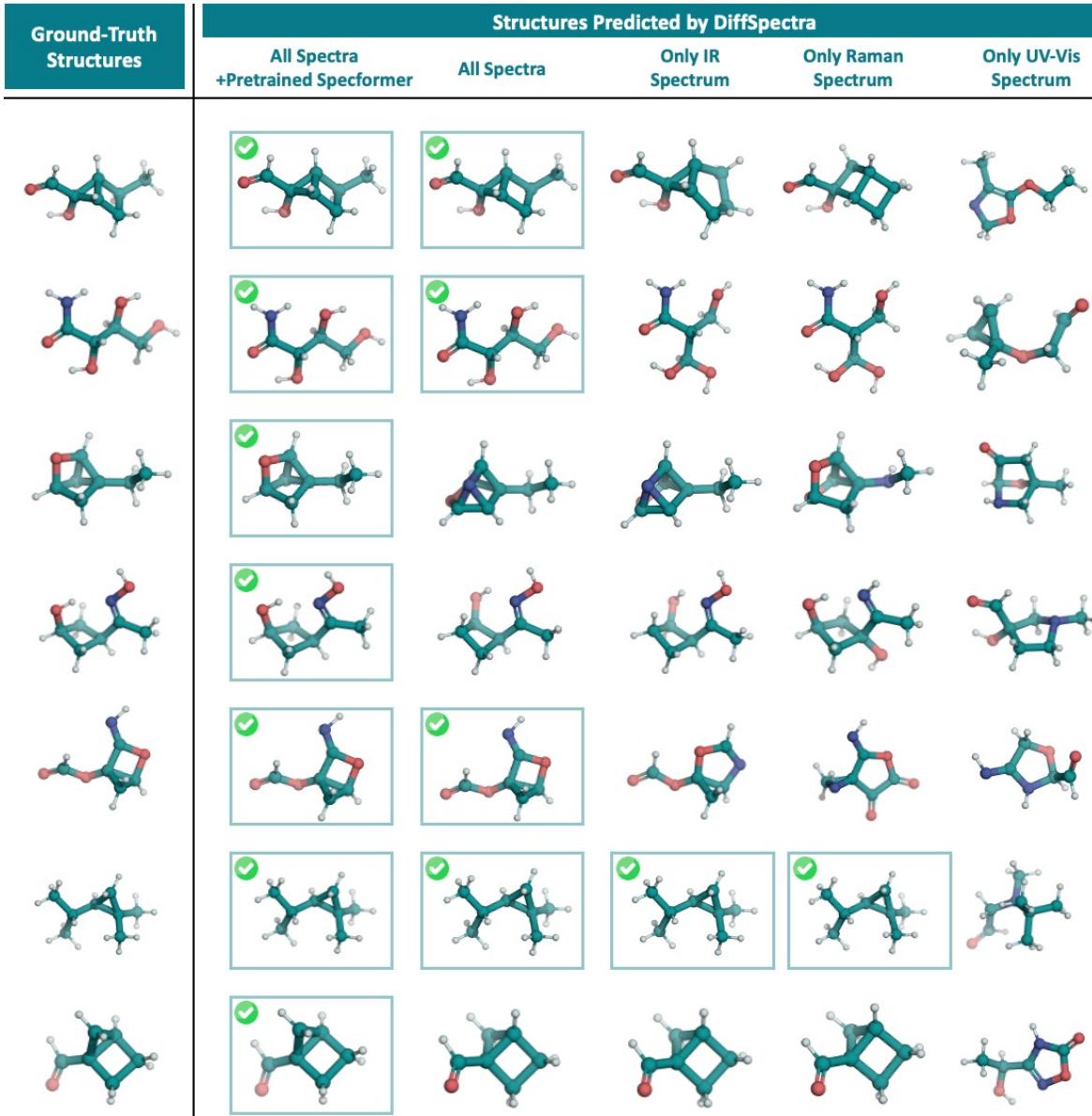
$$s_i = \max (\text{TaniSim}_{\text{RDKit}}(\mathbf{a}_{\text{RDKit}}, \hat{\mathbf{a}}_{\text{RDKit}}), \text{TaniSim}_{\text{RDKit}}(\mathbf{a}_{\text{RDKit,mask}}, \hat{\mathbf{a}}_{\text{RDKit,mask}}))$$

$$\text{FraggleSim} = \mathbb{E}_{\mathcal{G}} \left[\max_{f_i \in \mathcal{F}} s_i \right]$$

Functional Group-based Similarity (FGSim)

$$\text{FGSim} = \mathbb{E}_{\mathcal{G}} \left[\frac{|FG(\mathcal{G}) \cap FG(\hat{\mathcal{G}})|}{|FG(\mathcal{G}) \cup FG(\hat{\mathcal{G}})|} \right]$$

Experimental Results



1. DiffSpectra accurately elucidates molecular structures from spectra

Table 2: Structure elucidation performance of DiffSpectra on the QM9S dataset. We compare two configurations: one using a pre-trained SpecFormer as the spectral condition encoder, and one using an untrained SpecFormer. Reported metrics include exact structure recovery, graph overlap, fingerprint-based similarities, fragment-level similarity, and functional group similarity.

Model	Pre-trained SpecFormer	ACC@1 ↑	MCES ↓	TaniSim _{MG} ↑	CosSim _{MG} ↑	TaniSim _{MA} ↑	FraggleSim ↑	FGSim ↑
DiffSpectra	✓ ✗	16.01% 14.11%	1.3552 1.7795	0.7837 0.7205	0.8421 0.7938	0.9227 0.8924	0.9481 0.9383	0.9618 0.9490

2. Multi-modal spectra outperform single-modality spectra

Table 3: Effect of spectral modalities on structure elucidation performance. We report DiffSpectra's results on the QM9S dataset using different types of spectra as conditional input, including IR, Raman, UV-Vis, and their combination. Metrics include exact structure recovery, graph overlap, fingerprint-based similarities, fragment-level similarity, and functional group similarity.

Model	Spectral Modalities	ACC@1 ↑	MCES ↓	TaniSim _{MG} ↑	CosSim _{MG} ↑	TaniSim _{MA} ↑	FraggleSim ↑	FGSim ↑
DiffSpectra	All Spectra	14.11%	1.7795	0.7205	0.7938	0.8924	0.9383	0.9490
	Only IR	10.97%	2.4812	0.6246	0.7188	0.8460	0.9197	0.9269
	Only Raman	12.51%	2.1708	0.6778	0.7612	0.8701	0.9343	0.9315
	Only UV-Vis	0.10%	8.7909	0.1556	0.2625	0.3634	0.5581	0.4567

Experimental Results

3. Sampling multiple candidates improves structural hit accuracy

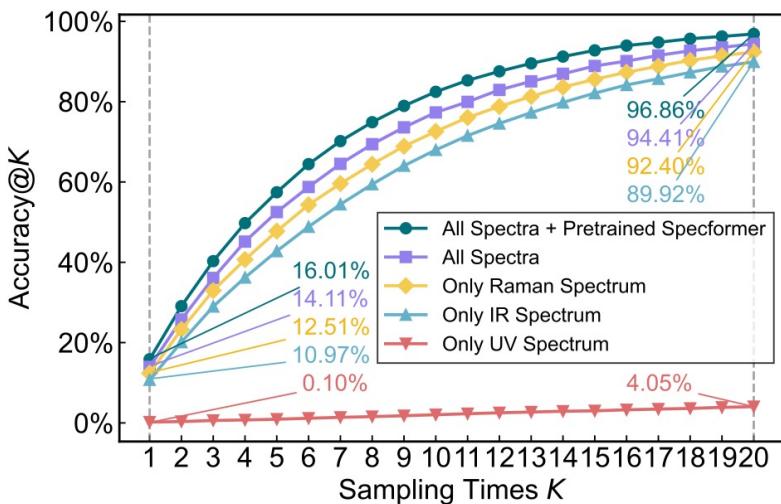


Fig. 3: Accuracy@ K with increasing number of sampled candidates. We evaluate the top- K accuracy (ACC@ K) as the number of generated candidates K increases, comparing different variants of our model. Across all settings, Accuracy@ K consistently improves with larger K , confirming that multiple sampling significantly increases the chance of recovering the correct molecular structure.

4. Comparing model-based and data-based SE(3) equivariance strategies

Table 4: Evaluation of different SE(3) equivariance strategies for DiffSpectra. We compare a model-based equivariant architecture to data-based equivariance approaches, with or without data augmentation, on molecular structure elucidation metrics. Here, “data aug” refers to data augmentation.

Model	SE(3) Equivariance	ACC@1 ↑	MCES ↓	TaniSim _{MG} ↑	CosSim _{MG} ↑	TaniSim _{MA} ↑	FraggleSim ↑	FGSim ↑
DiffSpectra	model-based	14.11%	1.7795	0.7205	0.7938	0.8924	0.9383	0.9490
	data-based (w/ data aug)	12.98%	1.8785	0.8882	0.7877	0.8882	0.9389	0.9456
	data-based (w/o data aug)	7.47%	3.6036	0.5117	0.6294	0.7575	0.8665	0.8607

5. Sampling temperature balances diversity and accuracy

$$\mathbf{G}_s = \frac{\alpha_{t|s}\sigma_s^2}{\sigma_t^2} \mathbf{G}_t + \frac{\alpha_s\sigma_{t|s}^2}{\sigma_t^2} \hat{\mathbf{G}}_0 + \tau \cdot \frac{\sigma_s\sigma_{t|s}}{\sigma_t} \mathbf{G}_e. \quad (39)$$

Table 5: Effect of sampling temperature on structure elucidation performance. We report DiffSpectra results with different sampling temperature coefficients τ , which scale the injected noise during diffusion sampling. Moderate values of τ help balance diversity and accuracy, while extremely low or high τ degrade performance.

Model	Temperature	ACC@1 ↑	MCES ↓	TaniSim _{MG} ↑	CosSim _{MG} ↑	TaniSim _{MA} ↑	FraggleSim ↑	FGSim ↑
DiffSpectra	$\tau=1.2$	15.45%	1.4224	0.7739	0.8348	0.9197	0.9468	0.9604
	$\tau=1.0$	16.01%	1.3552	0.7837	0.8421	0.9227	0.9481	0.9618
	$\tau=0.8$	16.30%	1.3643	0.7848	0.8429	0.9235	0.9491	0.9621
	$\tau=0.6$	16.05%	1.4009	0.7794	0.8387	0.9220	0.9480	0.9627
	$\tau=0.4$	15.86%	1.5081	0.7629	0.8261	0.9120	0.9442	0.9577
	$\tau=0.2$	14.56%	1.9842	0.7005	0.7774	0.8743	0.9297	0.9323



中国科学院自动化研究所
模式识别实验室
New Laboratory of Pattern Recognition



多模态人工智能系统
全国重点实验室
State Key Laboratory of
Multimodal Artificial Intelligence Systems



中国科学院自动化研究所
Institute of Automation
Chinese Academy of Sciences



NUS
National University
of Singapore

Thank you for your attention!

Contact : liang.wang@u.nus.edu